

11-05-07

AF



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:) Examiner: O'Hara, Eileen B.
)
Avi J. ASHKENAZI, et al.) Art Unit: 1646
)
Application Serial No. 09/978,192) Confirmation No: 3437
)
Filed: October 15, 2001) Attorney's Docket No. 39780-2630 P1C9
)
For: **SECRETED AND**) **Customer No. 35489**
 TRANSMEMBRANE)
 POLYPEPTIDES AND NUCLEIC)
 ACIDS ENCODING THE SAME)

EXPRESS MAIL LABEL NO. EV 582 638 189 US
DATE MAILED: November 1, 2007

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES
APPELLANTS' BRIEF

MAIL STOP APPEAL BRIEF - PATENTS

Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

11/06/2007 EHAILE1 00000042 001641 09978192
01 FC:1432 510.00 DA

Dear Sir:

On November 1, 2006, the Examiner made a final rejection to pending Claims 58-62. A Response to Final Office Action was filed on February 1, 2007. An Advisory Action was mailed on March 1, 2007. A Notice of Appeal was subsequently filed on April 2, 2007.

Appellants hereby appeal to the Board of Patent Appeals and Interferences from the last decision of the Examiner. A request for a five-month extension of time is requested with necessary fees.

The following constitutes Appellants' Brief on Appeal.

11/06/2007 EHAILE1 00000037 001641 09978192
01 FC:1255 2230.00 DA

11/06/2007 EHAILE1 00000037 001641 09978192
02 FC:1431 510.00 DA

1. REAL PARTY IN INTEREST

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application U.S. Patent Application Serial No. 09/918,585 recorded October 24, 2001, at Reel 012095 and Frame 0677.

2. RELATED APPEALS AND INTERFERENCES

The claims pending in the current application are directed to antibodies to a polypeptide referred to herein as "PRO274." There exists one related pending patent application, U.S. Patent Application Serial No. 09/978,188, filed October 15, 2001 (containing claims directed to PRO274 polypeptides). This related application is also under final rejection from the same Examiner and based upon very similar reasons, wherein appeal of these final rejections are being pursued independently and concurrently herewith. Although there exist several applications directed to the "gene amplification" utility, in general, under Appeal, none of these are related to PRO274 molecules or antibodies binding to it.

3. STATUS OF CLAIMS

Claims 58-62 are in this application.

Claims 1-57 and 63 are canceled.

Claims 58-62 stand rejected and Appellants appeal the rejection of these claims.

A copy of the rejected claims in the present Appeal is provided as Appendix A.

4. STATUS OF AMENDMENTS

All prior amendments have been entered by the Examiner. No claim amendments have been submitted after the Response filed August 3, 2006.

5. SUMMARY OF THE INVENTION

The invention claimed in the present application concerns an isolated antibody that specifically binds to the polypeptide of SEQ ID NO:7 (Claim 58). The invention further provides monoclonal antibodies (Claim 59), humanized antibodies (Claim 60), antibody fragments (Claim 61), and labeled antibodies (Claim 62) that specifically bind to the polypeptide of SEQ ID NO:7.

Support for the preparation and uses of antibodies is found throughout the specification, including, for example, pages 217-225. The preparation of antibodies is described in

Example 104, while Example 106 describes the use of the antibodies for purifying the polypeptides to which they bind. Isolated antibodies are defined in the specification at page 132, lines 29-38. Support for monoclonal antibodies is found in the specification at, for example, page 217, line 30, to page 219, line 11, and Example 104. Support for humanized antibodies is found in the specification at, for example, page 219, line 12, to page 220, line 14. Support for antibody fragments is found in the specification at, for example, page 131, line 29, to page 132, line 22, and page 221, lines 6-34. Support for labeled antibodies is found in the specification at, for example, page 133, lines 1-4, and page 224, line 35, to page 225, line 4.

The polypeptide of SEQ ID NO:7 is designated PRO274, and its amino acid sequence is shown in Figure 4, while the encoding nucleic acid sequence (SEQ ID NO:6) is shown in Figure 3. The specification discloses that various portions of the PRO274 polypeptide possess significant sequence similarity to the seven transmembrane receptor proteins (see, for example, page 2, line 27 to page 3, line 6). The isolation of cDNA clones encoding PRO274 of SEQ ID NO:7 is described in Example 4. Examples 100-103 describe the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells. Finally, Example 114, in the specification at page 331, line 23, to page 346, line 4, sets forth a Gene Amplification assay which shows that the PRO274 gene is amplified in the genome of certain human lung and colon cancers (see page Table 9).

The specification discloses that antibodies to PRO polypeptides may be used, for example, in purification of PRO (page 225, lines 5-11 and Example 106), in diagnostic assays for PRO expression (page 190, lines 3-9, and page 224, line 21 to page 225, line 4), as antagonists to PRO (page 198, lines 3-6), and as elements of pharmaceutical compositions for the treatment of various disorders (page 223, line 30, to page 224, line 28).

6. **GROUND OF REJECTION TO BE REVIEWED ON APPEAL**

- I. Whether Claims 58-62 satisfy the utility requirement of 35 U.S.C. §101.
- II. Whether Claims 58-62 satisfy the enablement requirement of 35 U.S.C. §112, first paragraph.
- III. Whether Claims 58-62 are patentable under 35 U.S.C. §102(b) over Ho *et al.*, Science, Vol. 289, pp 265-270.
- IV. Whether Claims 59-62 are patentable under 35 U.S.C. §103(a) over Ho *et al.*, in view of Janeway *et al.*

7. **ARGUMENT**

A. **Summary of the Arguments:**

Issue I: Utility

Appellants rely upon the gene amplification data of the PRO274 gene for patentable utility of the PRO274 polypeptide and the antibodies that bind it. This data is clearly disclosed in the instant specification in Example 114, which discloses that the gene encoding PRO274 showed significant amplification in primary lung tumors. Appellants submit that one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO274 gene, that the PRO274 polypeptide is concomitantly over expressed and has utility in the diagnosis of lung cancer or for individuals at risk for developing lung cancer.

The Examiner has asserted that it does not necessarily follow that an increase in gene copy number results in increased gene expression and increased protein expression, such that antibodies would be useful diagnostically. In support of these assertions, the Examiner referred to articles by Pennica *et al.* and Li *et al.* as evidence showing a lack of correlation between gene (DNA) amplification and mRNA levels, as well as articles by Gygi *et al.*, Hu *et al.*, Nagaraja *et al.*, Waghray *et al.*, and Sagynaliev *et al.* as providing evidence that polypeptide levels cannot be accurately predicted from mRNA levels.

Appellants submit that the teachings of the Examiner's cited references do not conclusively establish a *prima facie* case for lack of utility (as will be discussed in detail below).

As further support for their utility claim, Appellants have submitted a Declaration by Dr. Audrey Goddard, which explains that a gene identified as being amplified at least 2-fold by the disclosed

gene amplification assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy. Therefore, such a gene is useful as a marker for the diagnosis of lung cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy. According to the Goddard Declaration, the 2-fold to 3.1-fold amplification of PRO274 in three primary lung tumors would be considered significant and credible by one skilled in the art, based upon the facts disclosed therein. The Examiner has not provided any evidence to show that the disclosed DNA amplification is not significant.

Appellants have also submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level. For instance, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* collectively teach that in general, gene amplification increases mRNA expression. Of note, Appellants point out that the Examiner conceded to this argument in the preceding Advisory Action.

Further, Appellants have submitted over a hundred references, along with Declarations of Dr. Paul Polakis, which collectively teach that, in general, there is a correlation between mRNA levels and polypeptide levels. Appellants would also like to bring to the Examiner's attention a recent decision in a microarray case by the Board of Patent Appeals and Interferences (Decision on Appeal No. 2006-1469). In its decision, the Board reversed the utility rejection, acknowledging that "there is a strong correlation between mRNA levels and protein expression, and the Examiner has not presented any evidence specific to the PRO1866 polypeptide to refute that." (Page 9). Appellants submit that, in the instant application, the Examiner has likewise not presented any evidence specific to the PRO274 polypeptide to refute Appellant's assertion of a correlation between DNA levels, mRNA levels and protein expression.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is generally a positive correlation between DNA, mRNA, and polypeptide levels, in general, in the majority of amplified genes, as exemplified by the teachings of Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, the two Polakis Declarations, the art in general overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably

expect in this instance, based on the amplification data for the PRO274 gene, that the PRO274 polypeptide is concomitantly overexpressed and has utility in the diagnosis of lung cancer.

Appellants further submit that even if there were no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede to), a polypeptide encoded by a gene that is amplified in cancer would still have a specific, substantial, and credible utility. Appellants submit that, as evidenced by the Ashkenazi Declaration and the teachings of Hanna et al., simultaneous testing of gene amplification and gene product over-expression enables more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy for the tumor, as demonstrated by a real-world example of the breast cancer marker HER-2/neu.

Accordingly, Appellants submit that when the proper legal standard is applied, one should reach the conclusion that the present application discloses at least one patentable utility for the claimed antibodies to PRO274 polypeptides.

Issue II: Enablement

Appellants respectfully submit that the data presented in Example 114 of the specification and the cumulative evidence of record support a "specific, substantial and credible" asserted utility for the presently claimed invention. Accordingly, one of ordinary skill in the art would understand how to make and use the recited antibodies for the diagnosis of lung cancer without any undue experimentation.

These arguments are all discussed in further detail below under the appropriate headings.

Issue III: Anticipation by Ho et al.

Claims 58-62 stand rejected under 35 U.S.C. §102(b) as being anticipated by Ho et al., Science, Vol. 289, pp 265-270, published July 14, 2000.

The instant application claims priority to International Application No. PCT/US00/03565, which first disclosed the gene amplification results and was filed February 11, 2000, over five months before the publication date of Ho et al. The instant application has not been granted the earlier priority date on the grounds that "the gene amplification assay fails to disclose a patentable utility for the antibodies to the protein." (Page

10 of the Office Action mailed July 19, 2005). Appellants respectfully submit that as discussed above under Issues I and II, the presently claimed invention is supported by a specific, substantial and credible utility and, therefore, the present specification teaches one of ordinary skill in the art “how to use” the claimed invention without undue experimentation. Accordingly, the instant application is entitled to the effective filing date of February 11, 2000, and thus Ho et al. is not prior art.

Issue IV: Obviousness over Ho et al. in view of Janeway et al.

Claims 59-62 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Ho et al. in view of Immunology, The Immune System in Health and Disease, Third Edition, Janeway and Travers, Ed., 1997.

As discussed above, the instant application is entitled to priority to International Application No. PCT/US00/03565, and to the effective filing date of February 11, 2000. Thus Ho et al. is not prior art.

These arguments are all discussed in further detail below under the appropriate headings.

B. Detailed Arguments

Issue I: Claims 58-62 satisfy the utility requirement of 35 U.S.C. §101

The sole basis for the Examiner’s rejection of Claims 58-62 under these sections is that the data presented in Example 114 of the present specification is allegedly insufficient under applicable legal standards to establish a patentable utility under 35 U.S.C. §101 for the presently claimed subject matter.

Appellants strongly disagree and respectfully traverse the rejection.

i) The Legal Standard For Utility Under 35 U.S.C. §101

According to 35 U.S.C. §101:

Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

(Emphasis added).

In interpreting the utility requirement, in *Brenner v. Manson*,¹ the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent Appellant disclose a "substantial utility" for his or her invention, *i.e.*, a utility "where specific benefit exists in currently available form."² The Court concluded that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy."³

Later, in *Nelson v. Bowler*,⁴ the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The Court held that "since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility."⁵

In *Cross v. Iizuka*,⁶ the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that "*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results, *i.e.*, there is a reasonable correlation there between."⁷ The Court perceived, "No insurmountable difficulty" in finding that, under appropriate circumstances, "*in vitro* testing, may establish a practical utility."⁸

¹ *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

² *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

³ *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

⁴ *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

⁵ *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

⁶ *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

⁷ *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

⁸ *Id.*

The case law has also clearly established that Appellants' statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.⁹ The PTO has the initial burden to prove that Appellants' claims of usefulness are not believable on their face.¹⁰ In general, an Appellant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope."^{11, 12}

Compliance with 35 U.S.C. §101 is a question of fact.¹³ The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.¹⁴ Thus, to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Appellant. The issue will then be decided on the totality of evidence.

The well established case law is clearly reflected in the Utility Examination Guidelines ("Utility Guidelines"),¹⁵ which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted "specific, substantial, and credible utility" or a "well-established utility." Under the Utility Guidelines, a utility is "specific" when it is particular to the subject matter claimed. For example, it is generally not enough to state that a

⁹ *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

¹⁰ *Ibid.*

¹¹ *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

¹² See also *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (C.C.P.A. 1977).

¹³ *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) *cert. denied*, 469 US 835 (1984).

¹⁴ *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

¹⁵ 66 Fed. Reg. 1092 (2001).

nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the “substantial utility” standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an Appellant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a ‘substantial utility.’”¹⁶ Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement,¹⁷ gives the following instruction to patent examiners: “If the Appellant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility.”

ii) The Data and Documentary Evidence Supporting a Patentable Utility

Patentable utility for the PRO274 polypeptides and their antibodies is based upon the gene amplification data for the gene encoding the PRO274 polypeptide of SEQ ID NO: 7. Appellants respectfully submit that the data presented in Example 114 of the specification and the cumulative evidence of record support a “specific, substantial and credible” asserted utility for the presently claimed invention.

Example 114 describes the results obtained using a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqManTM PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9, including primary lung cancers of the type and stage indicated in Table 8 of the specification. The tumor samples were tested in

¹⁶ M.P.E.P. §2107.01.

¹⁷ M.P.E.P. §2107 II(B)(1).

triplicates with Taqman™ primers and with internal controls, beta-actin and GADPH in order to quantitatively compare DNA levels between samples. As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control and also, no-template controls. Gene amplification was monitored using real-time quantitative TaqMan™ PCR. Table 9 shows the resulting gene amplification data. Further, Example 114 explains that the results of TaqMan™ PCR are reported in Δ Ct units, wherein **one unit** corresponds to one PCR cycle or approximately a **2-fold amplification**, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on" (emphasis added). Appellants respectfully submit that a Δ Ct value of at least 1.0, which is a **more than 2-fold increase**, was observed for PRO274 in primary lung tumors LT4, LT16, and LT18. PRO274 showed approximately 1.00-1.61 Δ Ct units which corresponds to $2^{1.00}$ - $2^{1.61}$ fold, or 2.0-3.1 fold amplification in three different human primary lung tumors, which is significant and thus the PRO274 gene has utility as a diagnostic marker of lung cancer.

It is also well known that gene amplification occurs in most solid tumors, and generally is associated with poor prognosis.

In support, Appellants have submitted, in their Response filed September 14, 2004, a Declaration by Dr. Audrey Goddard. Appellants particularly draw the Board's attention to page 3 of the Goddard Declaration which clearly states that:

It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample **is significant** and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is **useful as a marker for the diagnosis of cancer**, for monitoring cancer development and/or for measuring the efficacy of cancer therapy. (Emphasis added).

As indicated above, the gene encoding the PRO274 polypeptide shows significantly higher than a two fold amplification in three different lung tumors. In addition, the Goddard Declaration clearly establishes that the TaqMan real-time PCR method described in Example 114 has gained wide recognition for its versatility, sensitivity and accuracy, and is in extensive use for

the study of gene amplification. The facts disclosed in the Declaration also confirm that based upon the gene amplification results, one of ordinary skill would find it credible that PRO274 is a diagnostic marker of lung cancer.

Further, as discussed in detail below, Appellants have provided ample evidence in the form of articles from the art, like Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, and over a 100 references (see Evidence List items 13-144) and Declarations by experts in the field of oncology and gene expression, i.e.: the Declarations by Dr. Paul Polakis (I and II) and by Dr. Avi Ashkenazi, to show that, in general, if a gene is amplified in cancer, it is “more likely than not” that the encoded protein will also be expressed at an elevated level.

The Examiner has asserted that “the specification provides data showing a very small increase in DNA copy number, approximately 2-fold, in a few tumor samples for PRO274.” (Page 4 of the Office Action mailed July 19, 2005). The Examiner further asserts that “it was imperative to find evidence in the relevant scientific literature whether or not a small increase in DNA copy number would be considered by the skilled artisan to be predictive of increased mRNA and polypeptide levels.” (Page 4 of the Office Action mailed July 19, 2005).

Appellants respectfully submit that the Examiner seems to be applying a heightened utility standard in this instance, which is legally incorrect. Appellants have shown that the gene encoding PRO274 demonstrated significant amplification, from 2.0-3.1 fold, in three lung tumors. As explained in the Declaration of Dr. Audrey Goddard (submitted with the Response filed September 14, 2004):

It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample **is significant** and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. (Emphasis added).

By referring to the 2.0-fold to 3.1-fold amplification of the PRO274 gene in lung tumors as “very small” the Examiner appears to ignore the teachings within an expert's declaration without any basis, or without presenting any evidence to the contrary. Appellants respectfully draw the Examiner's attention to the Utility Examination Guidelines (Part IIB, 66 Fed. Reg. 1098 (2001)) which state that:

"Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered".

Thus, barring evidence to the contrary, Appellants maintain that the 2.2 to 3.1-fold amplification disclosed for the PRO274 gene is significant and forms the basis for the utility claimed herein.

The Examiner has further asserted that "[g]iven that PRO274 was amplified in only a very small number of tumors of the same type, the data do not support the implicit conclusion of the specification that PRO274 shows a positive correlation with lung cancer, much less that the levels of PRO274 would be diagnostic of such." (Page 6 of the Office Action mailed May 20, 2004).

Appellants emphasize that they have shown significant DNA amplification in three out of the lung tumor samples in Table 9, Example 114 of the instant specification. The fact that not all lung tumors tested positive in this study does not make the gene amplification data less significant. As any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even, with most tumors. For example, the article by Hanna and Mornin (submitted with the Response filed September 14, 2004), discloses that the known breast cancer marker HER-2/neu is "amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma" (page 1, col. 1). In fact, some tumor markers are useful for identifying rare malignancies. That is, the association of the tumor marker with a particular type of tumor lesion may be rare, or, the occurrence of that particular kind of tumor lesion itself may be rare. In either event, even these rare tumor markers which do not give a positive hit for most common tumors, have great value in tumor diagnosis, and consequently, in tumor prognosis. The skilled artisan would certainly know that such tumor markers are useful for better classification of tumors. Therefore, whether the PRO274 gene is amplified in three lung tumors or in all lung tumors is not relevant to its identification as a tumor marker, or its patentable utility. Rather, the fact that the amplification data for PRO274 is considered significant is what lends support to its usefulness as a tumor marker.

The Examiner has asserted that “[t]he data presented in the specification were not corrected for aneuploidy” and cites a reference by Sen *et al.* in support of the assertion that “[a] slight amplification of a gene does not necessarily mean overexpression in a cancer tissue, but can merely be an indication that the cancer tissue is aneuploid.” (Page 6 of the Office Action mailed May 20, 2005).

Appellants submit that it is known in the art that detection of gene amplification can be used for cancer diagnosis regardless of whether the increase in gene copy number results from intrachromosomal changes or from chromosomal aneuploidy. As explained by Dr. Ashkenazi in his Declaration (submitted with Appellants' Response filed September 14, 2004),

An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

Hence, Appellants submit that gene amplification of a gene, whether by aneuploidy or any other mechanism, is useful as a diagnostic marker.

The Examiner has asserted that “[o]ne skilled in the art would do further research to determine whether or not the PRO274 polypeptide levels increased significantly in the tumor samples. The requirement for such further research makes it clear that the asserted utility is not yet in currently available form, i.e., it is not substantial.” (Page 4 of the Office Action mailed July 19, 2005).

As discussed above, M.P.E.P. §2107.01 cautions Office personnel not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an Appellant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a ‘substantial’ utility.”¹⁸ Indeed, the

¹⁸ M.P.E.P. §2107.01.

Guidelines for Examination of Applications for Compliance With the Utility Requirement,¹⁹ gives the following instruction to patent examiners: "If the Appellant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility."

Appellants' position is based on the overwhelming evidence from gene amplification data disclosed in the specification which clearly indicate that the gene encoding PRO274 is significantly amplified in certain lung tumors. Based on the working hypothesis among those skilled in the art that if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level, one skilled in the art would simply accept that since the PRO274 gene is amplified, the PRO274 polypeptide would be more likely than not over-expressed. Thus, data relating to PRO274 polypeptide expression may be used for the same diagnostic and prognostic purposes as data relating to PRO274 gene expression. Therefore, based on the disclosure in the specification, no further research would be necessary to determine how to use the claimed antibodies that bind to the PRO274 polypeptide, because the current invention is fully enabled by the disclosure of the present application.

Accordingly, Appellants submit that based on the general knowledge in the art at the time the invention was made and the teachings in the specification, the specification provides clear guidance as to how to interpret and use the data relating to PRO274 polypeptide expression and that the claimed antibodies which bind the PRO274 polypeptide have utility in the diagnosis of cancer.

iii) **A prima facie case of lack of utility has not been established**

Appellants submit that the evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Thus, to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner has

¹⁹ M.P.E.P. §2107 II(B)(1).

made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Appellant.

The Examiner has asserted that “it does not necessarily follow that an increase in gene copy number results in increased gene expression and increased protein expression, such that antibodies would be useful diagnostically.” (Page 7 of the Office Action mailed May 20, 2004). In support of these assertions, the Examiner referred to Pennica *et al.* and contended that “Pennica *et al.* was cited as evidence showing a lack of correlation between gene (DNA) amplification and mRNA levels.” (Page 4 of the Office Action mailed July 19, 2005). The Examiner further referred to Gygi *et al.*, and asserted that “Gygi *et al.* was cited as providing evidence that polypeptide levels cannot be accurately predicted from mRNA levels.” (Page 4 of the Office Action mailed July 19, 2005).

As a preliminary matter, Appellants respectfully submit that it is not a legal requirement to establish that gene amplification “necessarily” results in increased expression at the mRNA and polypeptide levels, or that protein levels can be “accurately predicted.” As discussed above, the evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Accordingly, Appellants submit that in order to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that **it is more likely than not** that one of ordinary skill in the art would doubt the truth of the statement of utility. Therefore, it is not legally required that there be a “necessary” correlation between the data presented and the claimed subject matter. The law requires only that one skilled in the art should accept that such a correlation is **more likely than not to exist.** Appellants respectfully submit that when the proper evidentiary standard is applied, a correlation must be acknowledged.

Pennica *et al.*

The Examiner cited the abstract of Pennica *et al.* for its disclosure that “WISP-1 gene amplification and overexpression in human colon tumors showed a correlation between DNA amplification and over-expression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with expression

in normal colonic mucosa from the same patient.” From this, the Examiner correctly concluded that increased copy number does not *necessarily* result in increased polypeptide expression. The standard, however, is not absolute certainty.

In fact, as noted even in Pennica *et al.*, “[a]n analysis of *WISP*-1 gene amplification and expression in human colon tumors *showed a correlation between DNA amplification and over-expression...*” (Pennica *et al.*, page 14722, left column, first full paragraph, emphasis added). Thus the findings of Pennica *et al.* with respect to *WISP*-1 support Appellants’ arguments. In the case of *WISP*-3, the authors report that there was no change in the DNA copy number, but there was a change in mRNA levels. This apparent lack of correlation between DNA and mRNA levels is not contrary to Appellants’ assertion that a change in DNA copy number generally leads to a change in mRNA level. Appellants are not attempting to predict the DNA copy number based on changes in mRNA level, and Appellants have not asserted that the only means for changing the level of mRNA is to change the DNA copy number. Therefore a change in mRNA without a change in DNA copy number is not contrary to Appellants’ assertions.

The fact that the single *WISP*-2 gene did not show the expected correlation of gene amplification with the level of mRNA/protein expression does not establish that it is more likely than not, in general, that such correlation does not exist. The Examiner has not shown whether the lack of correlation observed for the *WISP*-2 gene is typical, or is merely a discrepancy, an exception to the rule of correlation. Indeed, the working hypothesis among those skilled in the art is that, if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level, as was demonstrated for *WISP*-1.

Accordingly, Appellants respectfully submit that Pennica *et al.* teaches nothing conclusive regarding the absence of correlation between amplification of a gene and over-expression of the encoded *WISP* polypeptide. More importantly, the teaching of Pennica *et al.* is specific to *WISP* genes. Pennica *et al.* has no teaching whatsoever about the correlation of gene amplification and protein expression in general.

Gygi et al.

The Examiner further cited the Gygi *et al.* reference to establish that “the protein levels cannot be accurately predicted from the level of the corresponding mRNA transcript.” The

Examiner adds that “Gygi *et al.* ... studied over 150 proteins... and found no strong correlation between proteins and transcript levels.” (Page 7 of the Office Action mailed May 20, 2004).

Appellants respectfully traverse and point out that, on the contrary, Gygi *et al.* never indicate that the correlation between mRNA and protein levels does not exist. Gygi *et al.* only state that the correlation may not be sufficient in **accurately** predicting protein level from the level of the corresponding mRNA transcript (Emphasis added) (see page 1270, Abstract). This result is expected, since there are many factors that determine translation efficiency for a given transcript, or the half-life of the encoded protein. Not surprisingly, Gygi *et al.* concluded that protein levels cannot always be accurately predicted from the level of the corresponding mRNA transcript in a single cellular stage or type when looking at the level of transcripts across different genes.

Importantly, Gygi *et al.* did not say that for a single gene, a change in the level of mRNA transcript is not positively correlated with a change in the level of protein expression. Appellants have asserted that increasing the level of mRNA for a particular gene leads to a corresponding increase for the encoded protein. Gygi *et al.* did not study this issue and says absolutely nothing about it. One cannot look at the level of mRNA across several different genes to investigate whether a change in the level of mRNA for a particular gene leads to a change in the level of protein for that gene. Therefore, Gygi *et al.* is not inconsistent with or contradictory to the utility of the instant claims, and offers no support for the PTO’s rejection of Appellants’ asserted utility.

Furthermore, Appellants note that contrary to the Examiner's statement, the Gygi data indicate **a general trend** of correlation between protein [expression] and transcript levels (Emphasis added). For example, as shown in Figure 5, an mRNA abundance of **250-300** copies /cell correlates with a protein abundance of **500-1000** x 10³ copies/cell. An mRNA abundance of **100-200** copies/cell correlates with a protein abundance of **250-500** x 10³ copies/cell (emphasis added). Therefore, high levels of mRNA **generally** correlate with high levels of proteins. In fact, most data points in Figure 5 did not deviate or scatter away from the general trend of correlation. Thus, the Gygi data meets the “more likely than not standard” and shows that a positive correlation exists between mRNA and protein. Therefore, Appellants submit that the Examiner's rejection is based on a misrepresentation of the scientific data presented in Gygi *et al.*

Gygi *et al.* may teach that protein levels cannot be “accurately predicted” from mRNA levels in the sense that the exact numerical amounts of protein present in a tissue cannot be determined based upon mRNA levels. Appellants respectfully submit that the Office Action’s emphasis on the need to “accurately predict” protein levels based on mRNA levels misses the point. The asserted utility for the claimed polypeptides is in the diagnosis of cancer. What is relevant to use as a cancer diagnostic is relative levels of gene or protein expression, not absolute values, that is, that the gene or protein is differentially expressed in tumors as compared to normal tissues. Appellants need only show that there is a correlation between DNA, mRNA, and protein levels, such that gene amplification and mRNA overexpression generally predict protein overexpression. A showing that mRNA levels can be used to “accurately predict” the precise levels of protein expression is not required.

In conclusion, the Examiner has not shown that a lack of correlation between gene amplification: polypeptide over-expression, as observed for the *WISP-2* gene, is typical. In fact, contrary to what the Examiner contends, the art indicates that, if a gene is amplified in cancer, it is **more likely than not** that the encoded protein will be expressed at an elevated level. As noted even in Pennica *et al.*, a correlation between DNA amplification: polypeptide over-expression was observed in the case of *WISP-1* and similarly, in Gygi *et al.*, **most genes** showed a correlation between mRNA levels and protein levels. Since the standard is not absolute certainty, a *prima facie* showing of lack of utility has not been made in this instance.

Hu *et al.*

The Examiner further cited Hu *et al.* to the effect that genes displaying a 5-fold change or less in mRNA expression in tumors compared to normal showed no evidence of a correlation between altered gene expression and a known role in the disease. However, among genes with a 10-fold or more change in expression level, there was a strong and significant correlation between expression level and a published role in the disease. (Page7 of the Office Action mailed July 19, 2005).

Appellants first note that the title of Hu *et al.* is “Analysis of Genomic and Proteomic Data Using Advanced Literature Mining.” As the title clearly suggests, the conclusion suggested by Hu *et al.* is merely based on a statistical analysis of the information disclosed in the published

literature. As Hu *et al.* states, “We have utilized a computational approach to literature mining to produce a comprehensive set of gene-disease relationships.” In particular, Hu *et al.* relied on the MedGene Database and the Medical Subject Heading (MeSH) files to analyze the gene-disease relationship. More specifically, Hu *et al.* “compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples.” (See page 408, right column). Therefore, Appellants first submit that the reference by Hu *et al.* only studies the statistical analysis of micro-array data and not gene amplification data. Thus their findings would not be directly applicable to gene amplification data.

According to Hu *et al.*, “different statistical methods” were applied to “*estimate* the strength of gene-disease relationships and evaluated the results.” (See page 406, left column, emphasis added). Using these different statistical methods, Hu *et al.* “[a]ssessed the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation.” (See page 411, left column). It is well known in the art that various statistical methods allow different variables to be manipulated to affect the outcome. For example, the authors admit, “Initial attempts to search the literature using” the list of genes, gene names, gene symbols, and frequently used synonyms, generated by the authors “revealed several sources of false positives and false negatives.” (See page 406, right column). The authors further admit that the false positives caused by “duplicative and unrelated meanings for the term” were “difficult to manage.” Therefore, in order to minimize such false positives, Hu *et al.* disclose that these terms “had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.” *Id.* Hence, Appellants respectfully submit that in order to minimize the false positives and negatives in their analysis, Hu *et al.* manipulated various aspects of the input data.

Appellants further submit that the statistical analysis by Hu *et al.* is not a reliable standard because the frequency of citation reflects only the current research interest of a molecule rather than the true biological function of the molecule. Indeed, the authors acknowledge that “[r]elationship established by frequency of co-citation do not necessarily represent a true biological link.” (See page 411, right column). One would expect that genes with the greatest change in expression in a disease would be the first targets of research, and therefore have the

strongest known relationship to the disease as measured by the number of publications reporting a connection with the disease. The correlation reported in Hu only indicates that the greater the change in expression level, the more likely it is that there is a published or known role for the gene in the disease, as found by their automated literature-mining software. Thus, Hu's results merely reflect a bias in the literature toward studying the most prominent targets, and say nothing regarding the ability of a gene that is 2-fold or more differentially expressed in tumors to serve as a disease marker.

Even assuming that Hu *et al.* provide evidence to support a true relationship, the conclusion in Hu *et al.* only applies to a specific type of breast tumor (estrogen receptor (ER)-positive breast tumor) and can not be generalized as a principle governing microarray study of breast cancer in general, let alone the various other types of cancer genes in general. In fact, even Hu *et al.* admit that, "[i]t is likely that this threshold will change depending on the disease as well as the experiment. Interestingly, the observed correlation was only found among ER-positive (breast) tumors not ER-negative tumors." (See page 412, left column). Therefore, based on these findings, the authors add, "This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently." *Id.* (Emphasis added).

More importantly, Hu *et al.* did not look for a correlation between changes in mRNA and changes in protein levels, and therefore their results are not contrary to Appellants' assertion that there is a correlation between the two. Appellants are not relying on any "biological role" that the PRO274 polypeptide has in cancer for its asserted utility. Instead, Appellants are relying on the amplification of the gene encoding PRO274 in certain tumors compared to their normal tissue counterparts. Nowhere in Hu does it say that a lack of correlation in their study means that genes with a less than five-fold change in level of expression in cancer cannot serve as a diagnostic marker of cancer.

Li et al.

The Examiner also cites Li *et al.* as teaching that "68.8% of the genes showing over-representation in the genome did not show elevated transcript levels." (Page 13 of the Office

Action mailed November 1, 2006). Appellants respectfully point out that Li *et al.* acknowledge that their results differed from those obtained by Hyman *et al.* and Pollack *et al.* (of record), who found a substantially higher level of correlation between gene amplification and increased gene expression. The authors note that “[t]his discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma” (page 2629, col. 1). In fact, as explained in the Supplemental Information accompanying the Li article, genes were considered to be amplified if they had a copy number ratio of at least 1.40. As discussed in Appellants’ previous responses, and in the Goddard Declaration of record, an appropriate threshold for considering gene amplification to be significant is a copy number of at least 2.0. As discussed above the PRO274 gene showed 2.0 fold to 3.05-fold amplification in three different lung tumors, thus meeting this standard. It is not surprising that, by using a substantially lower threshold for considering a gene to be amplified, Li *et al.* would have identified a number of genes that were not in fact significantly amplified, and therefore did not show any corresponding increase in mRNA expression. The results of Li *et al.* therefore do not disprove that a gene with a substantially higher level of gene amplification, such as PRO274, would be expected to show a corresponding increase in transcript expression. Appellants point out that the Examiner has found these arguments persuasive as noted in the Advisory Action mailed March 1, 2007.

Nagaraja *et al.*, Waghray *et al.*, and Sagynaliev *et al.*

In further support of the assertion that “*changes* in mRNA expression frequently do not result in *changes* in protein expression” (page 16 of the Office Action mailed November 1, 2006; emphasis in original), the Examiner cites three additional references, by Nagaraja *et al.*, Waghray *et al.*, and Sagynaliev *et al.*

The Examiner cites Nagaraja *et al.* as allegedly teaching that in comparisons of expression profiles for normal breast compared to breast cancer, “the proteomic profiles indicated altered abundance of fewer proteins as compared to transcript profiles.” (Page 15 of the Office Action mailed November 1, 2006).

Appellants respectfully submit that the fact that many more transcripts than proteins were found to be differentially expressed does not mean that most mRNA changes did not result in

correlating protein changes, but merely reflects the fact that expression levels were only measured at all for many fewer proteins than transcripts. In particular, the total number of proteins whose expression levels could be visualized on silver-stained gels was only about 300 (page 2332, col. 1), as compared to the approximately 14,500 genes on the microarray chips for which mRNA levels were measured (page 2336, col. 1). Since the expression levels of so many fewer proteins than transcripts were measured, it is hardly surprising that a smaller absolute number of proteins than mRNAs were found to be overexpressed, because the protein products of most of the overexpressed mRNAs would not have been among the small number of proteins identified on the gels.

The Examiner next cites Waghray *et al.*, to the effect that “for most of the proteins identified, there was no appreciable concordant change at the RNA level.” (Page 15 of the Office Action mailed November 1, 2006).

Appellants emphasize that Appellants are asserting that a measurable change in mRNA level generally leads to a corresponding change in the level of protein expression, not that changes in protein level can be used to predict changes in mRNA level. Waghray et al. did not take genes which showed significant mRNA changes and check the corresponding protein levels. Instead, the authors looked at a small and unrepresentative number of proteins, and checked the corresponding mRNA levels. Waghray *et al.* acknowledge that only “[a] relatively small set of genes could be analyzed at the protein level, largely due to the limited sensitivity of 2-D PAGE” (page 1337, col. 1). In particular, while the authors examined the expression levels of 16,570 genes (page 1329, col. 2), they were able to measure the expression levels of only 1031 proteins (page 1333, col. 2). Waghray *et al.* does not teach that changes in mRNA expression were not correlated with changes in expression of the corresponding protein. All Waghray *et al.* state is that “for most of the proteins identified, there was no appreciable concordant change at the mRNA level” (page 1337, col. 2). This statement is not relevant to Appellants’ assertion of utility, since Appellants are not asserting that changes in mRNA levels are the only cause of changes in protein levels. Waghray *et al.* do not contradict Appellants’ assertion that changes in mRNA expression, in general, correspond to changes in expression of the corresponding protein. Of note, the Examiner has withdrawn the rejection based on Waghray *et al.* in the Advisory Action of March 1, 2007.

Lastly, the Examiner cites Sagynaliev *et al.*, as allegedly teaching that “it is also difficult to reproduce transcriptomics results with proteomics tools.” In particular, the Examiner notes that according to Sagynaliev *et al.*, of 982 genes found to be differentially expressed in *Human CRC*, only 177 (18%) have been confirmed using proteomics technologies. (Page 16 of the Office Action mailed November 1, 2006).

The Sagynaliev *et al.* reference, titled “Web-based data warehouse on gene expression in *Human* colorectal cancer” (emphasis added), drew conclusions based upon a literature survey of gene expression data published in *Human CRC*, and not from experimental data. While a literature survey can be a useful tool to assist researchers, the results may greatly over-represent or under-represent certain genes, and thus the conclusions may not be generally applicable. In particular, Appellants note that, as evidenced by Nagaraja *et al.* and Waghray *et al.*, discussed above, the number of mRNAs examined in transcriptomics studies is typically much larger than the number of proteins examined in corresponding proteomics studies, due to the difficulties in detecting and resolving more than a small minority of all expressed proteins on 2D gels. Thus, the fact that only 18% of all genes found to be differentially expressed in *Human CRC* have been confirmed using proteomics technologies does not mean that the corresponding proteins are not also differentially expressed, but is most likely due to the fact that the corresponding proteins were not identified on 2D gels, and thus their expression levels remain unknown.

The authors of Sagynaliev *et al.* acknowledge the many technical problems in finding proteomic data for CRC that can be matched to transcriptomic data to see if the two correlate. The authors state that “results have been obtained using heterogeneous samples in particular cell lines, whole tissue biopsies, and epithelial cells purified from surgical specimens.” However, “Results obtained in cell lines do not allow accurate comparison between normal and cancer cells, and the presence/absence of proteins of interest has to be confirmed in biopsies.” (Page 3072, left column.) In particular, the authors specifically note that “only a single study [1] provided differential display protein expression data obtained in the *Human* patient, using whole tissue biopsy.” (Page 3068, left column, second paragraph; *see also*, Table 2.) The examiner also notes and the authors state, “For CRC, there is no publication comparing mRNA and protein expression for a cohort of genes.” (Page 3077, left column, last paragraph, emphasis added.)

Appellants further note that Table 2 shows that 6 out of 8 published proteomics studies were

done using 2-D PAGE. However, the authors state that “2-D PAGE or 2-D DIGE have well-known technological limitations ... even under well-defined experimental conditions, 2-D PAGE parallel analysis of paired CRC samples is hampered by a significant variability.” (Page 3077, left column, third paragraph.) Therefore, Appellants respectfully submit that it is well known in the art that there are problems associated with selecting only those proteins detectable by 2D gels.

The Examiner further asserts that “the specification of the instant application does not teach a change in mRNA level of PRO274” because “[t]here are no teachings in the specification as to the differential expression of PRO274 mRNA in the progression of lung cancer or in response to different treatments of hormones (for example).” (Page 16 of the Office Action mailed November 1, 2006). Appellants respectfully note that the instant specification measured gene amplification, not mRNA expression. Appellants further submit that it is well known that cancers arise from the transformation of normal tissue cells to cancerous cells, *thus* the observed differences in gene amplification between normal and cancerous tissues are in fact the result of previously occurring changes.

Finally, in the Advisory Action of March 1, 2007, the Examiner notes that Sagynaliev points out that many genes found to be differentially regulated do not play a causal role in CRC carcinogenesis.

The Examiner appears to be concerned with the underlying mechanism resulting in the positive gene amplification results, and not with those results themselves. However, the Examiner’s concerns regarding the causal role of PRO274 associated with any type of cancer versus normal tissue, in no way negate the utility of the claimed invention. Appellants are relying on diagnostic utility and not therapeutic utility. The fact remains that the gene amplification results demonstrate overexpression of PRO274 in the named tumor. One of ordinary skilled in the art does not need to know the underlying mechanism of the overexpression of PRO274, or the downstream effects of that overexpression, to practice the diagnostic utility. One of ordinary skill in the art, in possession of these results, would have believed it more likely than not that the PRO274 polypeptide and the antibodies that binds it were useful for their asserted utility.

In summary, Appellants respectfully submit that the Examiner has not shown that a change in gene expression level in tumor as compared to normal tissue is not correlated with a

change in protein expression. The Patent Office has failed to meet its initial burden of proof that Appellants' claims of utility are not substantial or credible. The arguments presented by the Examiner in combination with the Pennica *et al.*, Gygi *et al.* and Hu *et al.* articles, as well as those by Li *et al.*, Nagaraja *et al.*, Waghray *et al.*, and Sagynaliev *et al.*, do not provide sufficient reasons to doubt the statements by Appellants that PRO274 has utility. As discussed above, the law does not require that gene amplification "necessarily" results in increased expression at the mRNA and polypeptide levels. Therefore, Appellants submit that the Examiner's reasoning is based on a misrepresentation of the scientific data presented in the above cited references and application of an improper, heightened legal standard. In fact, contrary to what the Examiner contends, the art indicates that if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level.

iv) **It is "more likely than not" for amplified genes to have increased mRNA and protein levels**

Appellants have submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level. First, the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.*, (made of record in Appellants' Response filed September 14, 2004) collectively teach that in general, gene amplification increases mRNA expression.

Second, Appellants have submitted over a *Hundred* references, along with Declarations of Dr. Paul Polakis with their Preliminary Amendment filed on August 3, 2006, which collectively teach that, in general, there is a correlation between mRNA levels and polypeptide levels.

Third, Appellants would like to bring to the Examiner's attention a recent decision by the Board of Patent Appeals and Interferences (Decision on Appeal No. 2006-1469). In its decision, the Board reversed the utility rejection, acknowledging that "there is a strong correlation between mRNA levels and protein expression, and the Examiner has not presented any evidence specific to the PRO1866 polypeptide to refute that." (Page 9 of the Decision). Appellants submit that, in the instant application, the Examiner has likewise not presented any evidence specific to the PRO274 polypeptide to refute Appellants' assertion of a correlation between mRNA levels and protein expression.

Thus, taken together, all of the submitted evidence supports Appellants' position that gene amplification is more likely than not predictive of increased mRNA and polypeptide levels.

Appellants submit that there are numerous articles which show that generally, if a gene is amplified in cancer, it is more likely than not that the mRNA transcript will be expressed at an elevated level. For example, Orntoft *et al.* (*Mol. and Cell. Proteomics*, 2002, vol. 1, pages 37-45 - made of record in Appellants' Response filed September 14, 2004) studied transcript levels of 5600 genes in malignant bladder cancers, many of which were linked to the gain or loss of chromosomal material using an array-based method. Orntoft *et al.* showed that there was a gene dosage effect and taught that "in general (18 of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts" (see column 1, abstract). In addition, Hyman *et al.* (*Cancer Res.*, 2002, vol. 62, pages 6240-45 - made of record in Appellants' Response filed September 14, 2004) showed, using CGH analysis and cDNA microarrays which compared DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines, that there was "evidence of a prominent global influence of copy number changes on gene expression levels." (See page 6244, column 1, last paragraph). Additional supportive teachings were also provided by Pollack *et al.*, (*PNAS*, 2002, vol. 99, pages 12963-12968 - made of record in Appellants' Response filed September 14, 2004) who studied a series of primary Human breast tumors and showed that "...62% of highly amplified genes show moderately or highly elevated expression, and DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), and that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels." Thus, these articles collectively teach that in general, gene amplification increases mRNA expression.

Upon consideration of Appellants arguments made in the response of February 1, 2007, the Examiner states in the Advisory Action mailed March 1, 2007 that the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* "are convincing that gene amplification generally results in increased transcription."

In addition, in their Responses filed September 14, 2004 and August 3, 2006, Appellants submitted two Declaration by Dr. Polakis, principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, to show that mRNA expression

correlates well with protein levels, in general. As Dr. Polakis explains, the primary focus of the microarray project was to identify tumor cell markers useful as targets for both the diagnosis and treatment of cancer in *Humans*. The Declaration by Dr. Paul Polakis (Polakis I - made of record in Appellants' Response filed September 14, 2004) explains that in the course of Dr. Polakis' research using microarray analysis, he and his co-workers identified approximately 200 gene transcripts that are present in *Human* tumor cells at significantly higher levels than in corresponding normal *Human* cells. Appellants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels. The second Declaration by Dr. Polakis (Polakis II- - made of record in Appellants' Response filed August 3, 2006) presented evidentiary data in Exhibit B. Exhibit B of the Declaration identified 28 gene transcripts out of 31 gene transcripts (*i.e.*, greater than 90%) that showed good correlation between tumor mRNA and tumor protein levels. As Dr. Polakis' Declaration (Polakis II) says "[a]s such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA." Accordingly, Dr. Polakis has provided the facts to enable the Examiner to draw independent conclusions regarding protein data. Appellants further emphasize that the opinions expressed in the Polakis Declaration, including in the above quoted statement, are all based on factual findings. For instance, antibodies binding to about 30 of these tumor antigens were prepared and mRNA and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when *Human* tumor cells are compared with their corresponding normal cells. Therefore, Dr. Polakis' research, which is referenced in his Declaration, shows that, in general, there is a correlation between increased mRNA and polypeptide levels. Hence, one of skill in the art would reasonably expect that, based on the gene amplification data of the PRO274 gene, the PRO274 polypeptide is concomitantly overexpressed in lung tumors studied as well. Based on these experimental data and his vast scientific experience of more than 20 years, Dr. Polakis states that, for *Human* genes, increased mRNA levels typically correlate with an increase in abundance of the encoded protein. He further

confirms that “it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.”

Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip arrays in 2004. Clearly, the research community believes that the information obtained from these chips is useful (i.e., that it is more likely than not informative of the protein level).

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is a correlation between polypeptide and mRNA levels, these instances are exceptions rather than the rule. In the majority of amplified genes, the teachings in the art, as exemplified by Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, and the Polakis Declaration, overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Thus, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO274 gene, that the PRO274 polypeptide is concomitantly overexpressed. Accordingly, Appellants submit that the PRO274 polypeptides and antibodies have utility in the diagnosis of cancer and based on such a utility, one of skill in the art would know exactly how to use the claimed antibodies for diagnosis of cancer.

With respect to the correlation between mRNA expression and protein levels, the Examiner asserts that the Polakis Declaration (Polakis I) is insufficient to overcome the rejection of claims 58-62 since it is limited to a discussion of data regarding the correlation of mRNA levels and polypeptide levels and not gene amplification levels. The Examiner asserts that the Declaration does not provide data such that the Examiner can independently draw conclusions. (Page 7 of the Office Action mailed July 19, 2005).

Appellants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels, the correlation between gene amplification and mRNA levels having already been established by the data shown in the Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* articles. Appellants emphasize that the opinions expressed in the Polakis Declaration, including the quoted statement, are all based on factual findings. Subsequently, antibodies binding to about 30 of these tumor antigens were prepared, and mRNA

and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when Human tumor cells are compared with their corresponding normal cells. Dr. Polakis' statement that "an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell" is based on factual, experimental findings, clearly set forth in the Declaration. Accordingly, the Declaration is not merely conclusive, and the fact-based conclusions of Dr. Polakis would be considered reasonable and accurate by one skilled in the art.

The case law has clearly established that in considering affidavit evidence, the Examiner must consider all of the evidence of record anew.²⁰ "After evidence or argument is submitted by the Appellant in response, patentability is determined on the totality of the record, by a preponderance of the evidence with due consideration to persuasiveness of argument"²¹ Furthermore, the Federal Court of Appeals held in *In re Alton*, "[W]e are aware of no reason why opinion evidence relating to a fact issue should not be considered by an examiner."²² Appellants also respectfully draw the Examiner's attention to the Utility Examination Guidelines²³ which state, "Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered." The statement in question from an expert in the field (the Polakis Declaration) states that "it is my considered scientific opinion that for Human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell." Therefore, barring evidence to the contrary

²⁰ *In re Rinehart*, 531 F.2d 1084, 189 U.S.P.Q. 143 (C.C.P.A. 1976); *In re Piasecki*, 745 F.2d 1015, 226 U.S.P.Q. 881 (Fed. Cir. 1985).

²¹ *In re Alton*, 37 U.S.P.Q.2d 1578, 1584 (Fed. Cir. 1996)(quoting *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992)).

²² *Id.* at 1583.

²³ Part IIB, 66 Fed. Reg. 1098 (2001).

regarding the above statement in the Polakis Declaration, this rejection is improper under both the case law and the Utility guidelines.

The Examiner asserts that the second Polakis Declaration (Polakis II) is insufficient to overcome the utility rejection because PRO274 does not appear in the table (Exhibit B), and allegedly it is not clear whether PRO274 shares the same characteristics as those tumor antigens tested. (Page 5 of the Office Action mailed November 1, 2006).

Appellants respectfully submit that, as discussed in their previous Responses and Appeal Brief, the standard for utility is more likely than not. Dr. Polakis' Declarations provide evidence, in the form of statements by an expert in the art, that "an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell." The PRO274 gene was found to be amplified in lung tumors. As discussed above and in Appellants' previous Responses, one of ordinary skill in the art would therefore expect the PRO274 mRNA to be overexpressed in the same *Human* lung tumor samples. Accordingly, one of ordinary skill in the art would understand that the PRO274 polypeptide would be expected (more likely than not) to be overexpressed in *Human* lung tumor samples relative to their normal *Human* tissue counterparts; as are the majority of other molecules tested.

The Examiner further states that "levels of mRNA and protein in tumor tissue were compared to corresponding normal tissue, but the amplification levels of genomic DNA from example 114 were compared to normal *Human* blood, not corresponding normal tissue." (Page 5 of the Office Action mailed November 1, 2006). Appellants respectfully note that the Polakis Declaration describes the results of microarray experimentation, while Example 114 of the specification discloses gene amplification data. *Thus* the Examiner's attempt to contrast the methodology of the two types of experiments is misplaced.

The Examiner appears to require Appellants to provide every single experimental detail involved in the testing of the mRNA/protein correlation according to the Polakis Declaration. Such a requirement is unreasonable because neither the law nor the Utility Guidelines requires Appellants to do so.

The Examiner further notes (at page 6 of the Office Action mailed November 1, 2006) that Dr. Polakis is employed by the assignee. Appellants respectfully submit that note the sworn

Declaration of Dr. Polakis is sufficient to support Appellants' position a general mRNA/protein correlation, even if Dr. Polakis is an employee of the assignee.

Based on the above arguments, Appellants have clearly demonstrated a credible, specific and substantial asserted utility for the PRO274 polypeptide and the claimed antibodies that bind them, for example, as diagnostic markers for lung tumors. Further, based on this utility and the disclosure in the specification, one skilled in the art at the time the application was filed would know how to use the claimed antibodies.

With respect to the over one *Hundred* additional references supporting a correlation between mRNA and protein overexpression cited in Appellants' Preliminary Amendment filed August 3, 2006, the Examiner asserts that "[w]ith the exception of Futcher *et al.*, all of Appellant's newly cited references are directed to the analysis of single genes, or a small group of genes, and therefore do not demonstrate trends found across proteins in general." (Page 8 of the Office Action mailed November 1, 2006).

Appellants note that the submitted references, which represent experiments conducted by a large number of different research groups, demonstrate a trend of correlation found across proteins in general, and that this trend is confirmed by an overwhelming number of experiments by different researchers, using diverse experimental designs, testing various types of tissues, under numerous biological conditions. Although only a single gene or a small group of genes was tested by each individual study group, the cumulative evidence generated by over one *Hundred* study groups certainly establishes that it is well-accepted in the art that a general mRNA/protein correlation exists.

In response to the submitted textbook excerpts by Alberts and Lewin (submitted in the IDS filed on August 3, 2006 in Exhibit 1), the Examiner acknowledges that the teachings of Alberts and Lewin support that the initiation of transcription is the most common point for a cell to regulate gene expression. The Examiner asserts, however, that the initiation of transcription "is not the only means of regulating gene expression" according to the teaching of Alberts. (Page 7 of the Office Action mailed November 1, 2006).

Appellants respectfully submit that the utility standard is not **absolute certainty**. Rather, to overcome the presumption of truth that an assertion of utility by an Appellant enjoys, the PTO must establish that it is **more likely than not** that one of ordinary skill in the art would doubt the

truth of the statement of utility. Therefore, Appellants **do not need** to establish that transcription initiation is **the only means** of regulating gene expression in order to meet the utility standard. Instead, as long as it is the most common point of regulation, as admitted by the Examiner, it would be more likely than not that a change in the transcription level of a gene gives rise to a change in translation level of a gene. Appellants note that both Alberts and Lewin make clear that it is far more likely than not that protein levels for any given gene are regulated at the transcriptional level. Alberts, for example, states that of all the possible points for regulating protein expression, “[f]or most genes transcriptional controls are paramount.” Cell 4th at 379 (emphasis added). In a similar vein, Lewin states that “having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear that the overwhelming majority of regulatory events occur at the initiation of transcription.” *Genes VI* at 847-848 (emphasis added). Thus, the utility standard is met.

With respect to Appellants’ arguments regarding Meric *et al.* (submitted in the IDS filed on August 3, 2006 in Exhibit 1), the Examiner asserts that Meric teaches that “gene expression is quite complicated, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.” (Page 8 of the instant Office Action).

Appellants respectfully submit that Meric simply summarizes the translational regulation of cancer cells. Meric indicates that translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled to cell cycle progression and cell growth. Meric further discusses how alterations in translation control occur in cancer. For example, variant mRNA sequences can alter the translational efficiency of individual mRNA molecules. (see Abstract). Meric further teaches that the changes in translational efficiency of a mRNA transcript depend on the mutation of a specific mRNA sequence. (Page 973, column 2 to page 974, column 1). Meric never suggests that the translation of a cancer gene is suppressed in cancer in general, and that therefore, increased mRNA levels will not, in general, yield increased protein levels. To the contrary, Meric teaches that the translation efficiency of a number of cancer genes is enhanced in cancer cells compared to their normal counterparts. For instance, in patients with multiple myeloma, a C-T mutation in the c-myc IRES was identified and found to cause an enhanced initiation of translation (page 974, column 1). Therefore the level of proteins

encoded by these genes increases in cancer cells at an even higher magnitude than the corresponding mRNA level. *Thus* Meric clearly supports Appellants' assertions that it is more likely than not that, in general, changes in mRNA levels are correlated with changes in protein levels.

The Examiner further alleges that Meric teaches "translational efficiency of a number of cancer genes is enhanced in cancer cells compare to its normal counterpart due to mutation" and from this asserts that "the specification does not teach that PRO274 mRNA in cancer cells have a mutation that would lead to increased translation." (Advisory Action mailed March 1, 2007).

Appellants respectfully point out that Meric simply summarizes translational regulatory mechanisms in cancer cells and discusses how alterations in different aspects of translation control occur in cancer. Meric indicates that translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled to cell cycle progression and cell growth. Meric does not teach that mutation of a gene is required for its increased translation in cancer. Appellants do not need to establish the mechanism of regulating PRO274 gene expression in order to meet the utility standard.

The Examiner asserts that "the majority of the newly cited references by Appellants are drawn to genes known or suspected to be over expressed or under expressed in cancers, and that are involved with cell proliferation, differentiation and/or cell adhesion/migration, in which expression of the protein is important in the development and progression of the cancer." (Pages 8-9 of the Office Action mailed November 1, 2006).

Appellants respectfully submit that, in fact, a number of the references submitted with Appellants' IDS filed August 3, 2006, are drawn to proteins that are not members of the above protein categories and have no obvious association with cancer. To list just a few examples, Rudlowski *et al.* examined the expression of glucose transporters 1-4; Papotti *et al.* studied three somatostatin receptors; Van der Wilt *et al.* studied deoxycytidine kinase; and Grenback *et al.* studied galanin.

Appellants further respectfully submit that, as discussed in their previous Responses and Appeal Brief, Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.*, (made of record in Appellants' Response filed September 14, 2004), collectively teach that gene amplification increases mRNA expression for large numbers of genes, which have not been identified as being oncogenes or as

having any known functions in the development and progression of cancer. Thus the art of record clearly shows that there is no requirement that a polypeptide must be a known oncogene or a protein otherwise known to be associated with tumor growth, in order for amplification of the gene encoding the protein to correlate with increased protein expression. In fact, as demonstrated by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.*, examination of gene amplification is a useful way to identify novel proteins not previously known to be associated with cancer.

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is a correlation between polypeptide and mRNA levels, these instances are exceptions rather than the rule. In the majority of amplified genes, the teachings in the art, as exemplified by Orntoft *et al.*, Hyman *et al.*, Pollack *et al.*, and the Polakis Declarations, overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO274 gene, that the PRO274 polypeptide is concomitantly overexpressed. Thus, Appellants submit that the PRO274 polypeptide and the claimed antibodies that bind it have utility in the diagnosis of cancer.

v) **Even if a *prima facie* case of lack of utility has been established, it should be withdrawn on consideration of the totality of evidence**

Even if one assumes *arguendo* that it is more likely than not that there is no correlation between gene amplification and increased mRNA/protein expression, which Appellants submit is **not** true, a polypeptide encoded by a gene that is amplified in cancer would **still** have a specific, substantial, and credible utility. In support, Appellants respectfully draw the Board's attention to page 2 of the Declaration of Dr. Avi Ashkenazi (submitted with the Response filed September 14, 2004) which explains that,

even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-

expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

Appellants thus submit that simultaneous testing of gene amplification and gene product over-expression enables more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy. Further, as explained in Dr. Ashkenazi's Declaration, absence of over-expression of the protein itself is crucial information for the practicing clinician. If a gene is amplified in a tumor, but the corresponding gene product is not over-expressed, the clinician will decide not to treat a patient with agents that target that gene product. This not only saves money, but also has the benefit that the patient can avoid exposure to the side effects associated with such agents.

This utility is further supported by the teachings of the article by Hanna and Mornin. (Pathology Associates Medical Laboratories, August (1999); submitted with the Response filed September 14, 2004). The article teaches that the HER-2/neu gene has been shown to be amplified and/or over-expressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinomas. Further, the article teaches that diagnosis of breast cancer includes testing both the amplification of the HER-2/neu gene (by FISH) as well as the over-expression of the HER-2/neu gene product (by IHC). Even when the protein is not over-expressed, the assay relying on both tests leads to a more accurate classification of the cancer and a more effective treatment of it.

The Examiner asserts that "[t]he Hanna reference is not applicable to the instant fact situation, as it deals with a known tumor associated gene, and not with a prospective analysis of the type found in this specification." (Page 9 of the Office Action mailed July 19, 2005). To the contrary, Appellants have clearly shown that the gene encoding the PRO274 polypeptide is amplified in at least three primary lung tumors. Therefore, the PRO274 gene, similar to the HER-2/neu gene disclosed in Hanna *et al.*, is a tumor associated gene. Furthermore, as discussed above, in the majority of amplified genes, the teachings in the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO274 gene, that the PRO274 polypeptide is concomitantly overexpressed.

However, even if gene amplification does not result in overexpression of the gene product (*i.e.*, the protein) an analysis of the expression of the protein is useful in determining the course of treatment, as supported by the Ashkenazi Declaration. The Examiner has asserted that “the gene product of the instant invention has not been demonstrated to be involved in cancer. Overexpression of a gene product in a cancer cell does not necessarily mean that the gene product is involved in the cancer and that targeting the gene product would be therapeutic.” (Page 9 of the Office Action mailed July 19, 2005). The Examiner appears to view the testing described in the Ashkenazi Declaration and the Hanna paper as experiments involving further characterization of the PRO274 polypeptide itself. In fact, such testing is for the purpose of characterizing not the PRO274 polypeptide, but the tumors in which the gene encoding PRO274 is amplified. Testing of tumor markers such as PRO274 is useful for tumor categorization even if the tested marker is not itself the intended therapeutic target. The PRO274 polypeptide is therefore useful in tumor categorization, the results of which become an important tool in the hands of a physician enabling the selection of a treatment modality that holds the most promise for the successful treatment of a patient.

For the reasons given above, Appellants respectfully submit that the present specification clearly describes, details and provides a patentable utility for the claimed invention. Accordingly, Appellants respectfully request reconsideration and reversal of the rejections of Claims 58-62 under 35 U.S.C. §101.

Issue II: Claims 58-62 satisfy the enablement requirement of 35 U.S.C. §112, first paragraph.

Claims 58-62 stand rejected under 35 U.S.C. §112, first paragraph, allegedly “since the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention.” (Page 9 of the Office Action mailed July 19, 2005).

In this regard, Appellants refer to the arguments and information presented above in response to the outstanding rejection under 35 U.S.C. § 101, wherein those arguments are incorporated by reference herein. Appellants respectfully submit that as described above, the PRO274 polypeptide has utility in the diagnosis of cancer and based on such a utility, one of skill

in the art would know exactly how to use the claimed antibodies that bind the PRO274 polypeptide for diagnosis of cancer, without undue experimentation.

Accordingly, Appellants respectfully request reconsideration and reversal of the enablement rejection of Claims 58-62 under 35 U.S.C. §112, first paragraph.

Issue III: Claims 58-62 are patentable under 35 U.S.C. §102(b) over Ho *et al.*

Claims 58-62 stand rejected under 35 U.S.C. §102(b) as being anticipated by Ho *et al.*, Science, Vol. 289, pp 265-270, published July 14, 2000.

Appellants submit that, as discussed above in response to the outstanding rejections under 35 U.S.C. §101 and 35 U.S.C. §112, first paragraph, for alleged lack of utility and enablement Appellants rely on the gene amplification results (Example 114) to establish a credible, substantial and specific asserted utility for the PRO274 polypeptide and the claimed antibodies that bind it. These results were first disclosed in International Application No. PCT/US00/03565, filed February 11, 2000. As discussed above, the disclosure of the instant application, which is similar to that of the earlier-filed application (PCT/US00/03565), provides the support required under 35 U.S.C. §112 for the subject matter of the instant claims. Accordingly, Appellants submit that the subject matter of the instant claims is disclosed in the manner provided by 35 U.S.C. §112 in PCT/US00/03565. Therefore, the effective filing date of this application is February 11, 2000, the filing date of PCT/US00/03565.

The scientific journal article by Ho *et al.* was published on July 14, 2000, which is over five months after the effective filing date of the instant application; hence Ho *et al.* is not prior art.

The Examiner has asserted that the present claims are not entitled to the February 11, 2000, filing date of PCT/US00/03565 because “the gene amplification assay fails to disclose a patentable utility for the antibodies to the protein.” (Page 10 of the Office Action mailed July 19, 2005).

In this regard, Appellants refer to the arguments and information presented above in response to the outstanding rejections under 35 U.S.C. §101 and 35 U.S.C. §112, first paragraph, for alleged lack of utility and enablement. These arguments are incorporated by reference herein. Appellants respectfully submit that as described above under Issue I, the presently claimed

invention is supported by a specific, substantial and credible utility and, therefore, the present specification teaches one of ordinary skill in the art "how to use" the claimed invention without undue experimentation, as described above.

Accordingly, Appellants respectfully request reconsideration and reversal of the rejection of Claims 58-62 under 35 U.S.C. §102(b) as being anticipated by Ho *et al.*

Issue IV: Claims 59-62 are patentable under 35 U.S.C. §103(a) over Ho *et al.* in view of Janeway *et al.*

Claims 59-62 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Ho *et al.* in view of Immunology, The Immune System in Health and Disease, Third Edition, Janeway and Travers, Ed., 1997.

As discussed above, the effective filing date of this application is February 11, 2000, the filing date of PCT/US00/03565. The scientific journal article by Ho *et al.* was published on July 14, 2000, which is over five months after the effective filing date of the instant application; hence Ho *et al.* is not prior art, and is not available as a reference under 35 U.S.C. §103.

Accordingly, Appellants respectfully request reconsideration and reversal of the rejection of Claims 59-62 under 35 U.S.C. §103(a).

CONCLUSION

For the reasons given above, Appellants submit that the specification discloses at least one patentable utility for the antibodies of Claims 58-62, and that one of ordinary skill in the art would understand how to use the claimed antibodies, for example in the diagnosis of lung tumors. Therefore, Claims 58-62 meet the requirements of 35 U.S.C. §101 and 35 U.S.C. §112, first paragraph.

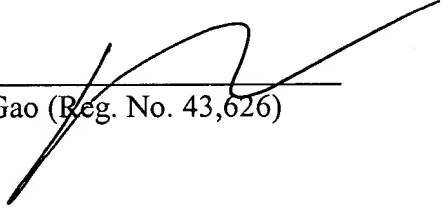
Further, this patentable utility for the claimed antibodies was first disclosed in International Application No. PCT/US00/03565, filed February 11, 2000, priority to which is claimed in the instant application. Accordingly, the instant application has an effective priority date of February 11, 2000, and therefore Ho *et al.*, Science, Vol. 289, pp 265-270, published on July 14, 2000, is not prior art and does not anticipate the claims under 35 U.S.C. §102(b) or render the claims obvious under 35 U.S.C. §103(a) in view of Janeway *et al.*

Accordingly, reversal of all the rejections of Claims 58-62 is respectfully requested.

Please charge any additional fees, including fees for additional extension of time, or credit overpayment to Deposit Account No. 08-1641 (referencing Attorney's Docket No. 39780-2630 P1C9).

Respectfully submitted,

Date: November 1, 2007

By: 
Panpan Gao (Reg. No. 43,626)

HELLER EHRMAN LLP
275 Middlefield Road
Menlo Park, California 94025-3506
Telephone: (650) 324-7000
Facsimile: (650) 324-0638
SV 2304999 v1
11/1/07 10:28 AM (39780.2630)

8. CLAIMS APPENDIX A

Claims on Appeal

- 58. An isolated antibody that specifically binds to the polypeptide of SEQ ID NO:7.
- 59. The antibody of Claim 58 which is a monoclonal antibody.
- 60. The antibody of Claim 58 which is a *Humanized* antibody.
- 61. An antigen binding fragment of the antibody of Claim 58.
- 62. The antibody of Claim 58 which is labeled.

9. EVIDENCE APPENDIX

1. Declaration of Audrey Goddard, Ph.D. under 35 C.F.R. §1.132, with attached Exhibits A-G:
 - A. Curriculum Vitae of Audrey D. Goddard, Ph.D.
 - B. Higuchi, R. *et al.*, "Simultaneous amplification and detection of specific DNA sequences," *Biotechnology* 10:413-417 (1992).
 - C. Livak, K.J., *et al.*, "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods Appl.* 4:357-362 (1995).
 - D. Heid, C.A. *et al.*, "Real time quantitative PCR," *Genome Res.* 6:986-994 (1996).
 - E. Pennica, D. *et al.*, "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in Human colon tumors," *Proc. Natl. Acad. Sci. USA* 95:14717-14722 (1998).
 - F. Pitti, R.M. *et al.*, "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer," *Nature* 396:699-703 (1998).
 - G. Bieche, I. *et al.*, "Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer," *Int. J. Cancer* 78:661-666 (1998).
2. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. §1.132, with attached Exhibit A (Curriculum Vitae).
3. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. §1.132 (Polakis I).
4. Orntoft, T.F., *et al.*, "Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-Invasive and Invasive Human Transitional Cell Carcinomas," *Molecular & Cellular Proteomics* 1:37-45 (2002).
5. Hyman, E., *et al.*, "Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer," *Cancer Research* 62:6240-6245 (2002).
6. Pollack, J.R., *et al.*, "Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors," *Proc. Natl. Acad. Sci. USA* 99:12963-12968 (2002).
7. Hanna, J.S., *et al.*, "HER-2/neu Breast Cancer Predictive Testing," Pathology Associates Medical Laboratories (1999).
8. Sen, S., "Aneuploidy and cancer," *Curr. Opin. Oncol.* 12:82-88 (2000).

9. Pennica, D. et al., "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in *Human colon tumors*," *Proc. Natl. Acad. Sci. USA* **95**:14717-14722 (1998).
10. Gygi, S. P. et al., "Correlation between protein and mRNA abundance in yeast," *Mol. Cell. Biol.* **19**:1720-1730 (1999).
11. Hu, Y. et al., "Analysis of genomic and proteomic data using advanced literature mining," *Journal of Proteome Research* **2**:405-412 (2003).
12. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. §1.132 (Polakis II).
13. Alberts, B., *et al.*, *Molecular Biology of the Cell* (3rd ed. 1994) Cell 3rd at 453 Figure 9-2 of Cell 3rd Cell 3rd at 403.
14. Alberts, B., *et al.*, *Molecular Biology of the Cell* (4rd ed.) In Cell 4th, Figure 6-3 on page 302 Figure 6-90 on page 364 of Cell 4th Cell 4th at 364 Cell 4th at 379.
15. Futcher, B., *et al.*, *Mol Cell Biol.*, - 19(11):7357-68 (1999).
16. Grenback, E., *et al.*, *Regul Pept*, - 117(2):127-39 (2004).
17. Lewin, B., *Genes VI* Genes VI at 847-848 (1997).
18. Meric, F., *et al.*, *Molecular Cancer Therapeutics* - 1:971-979 (2002).
19. Papotti, M., *et al.*, *Virchows Arch.* - 440(5):461-75 (2002).
20. Rudlowski, C., *et al.*, *Am J. Clin Pathol.* - 120(5):691-8 (2003).
21. Van Der Wilt, C.L., *et al.*, *Eur J Cancer* - 39(5):691-7 (2003).
22. Li *et al.*, 2006, *Oncogene*, 25: 2628-2635.
23. Nagaraja *et al.*, 2006, *Oncogene* 25: 2328-2338.
24. Waghray *et al.*, 2001, *Proteomics*, 1: 1327-1338.
25. Sagynaliev *et al.*, 2005, *Proteomics*, 5: 3066-3076.

Items 1-7 were submitted with Appellants' Response filed September 14, 2004, and acknowledged as having been considered by the Examiner in the Office Action mailed July 19, 2005.

Items 8-10 were made of record by the Examiner in the Office Action mailed May 20, 2004.

Item 11 was made of record by the Examiner in the Office Action mailed July 19, 2005.

Items 12-21 were submitted with Appellants' Preliminary Amendment filed August 3, 2006, and were considered by the Examiner as indicated in the Final Office Action mailed November 1, 2006.

Items 22-25 were made of record by the Examiner in the Final Office Action mailed November 1, 2006.

10. **RELATED PROCEEDINGS APPENDIX**

None.

SV 2304999 v1
10/31/07 1:33 PM (39780.2630)

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ashkenazi et al.	Group Art Unit: 1647
Serial No.: 09/903,925	Examiner: Fozia Hamid
Filed: July 11, 2001	CERTIFICATE OF MAILING I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on _____
For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS	Date _____

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.
2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

Serial No.: *

Filed: *

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl. 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: *

Filed: *

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003

Date

Audrey D. Goddard

Audrey D. Goddard, Ph.D.

AUDREY D. GODDARD, Ph.D.

Genentech, Inc.
1 DNA Way
South San Francisco, CA, 94080
650.225.6429
goddarda@gene.com

110 Congo St.
San Francisco, CA, 94131
415.841.9154
415.819.2247 (mobile)
agoddard@pacbell.net

PROFESSIONAL EXPERIENCE

Genentech, Inc.
South San Francisco, CA

1993-present

2001 - present Senior Clinical Scientist
Experimental Medicine / BioOncology, Medical Affairs

Responsibilities:

- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

Interests:

- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

1998 - 2001 Senior Scientist
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities:

- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a \$750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

Research:

- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

1993 - 1998 Scientist

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities

- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

Research:

- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

**Imperial Cancer Research Fund
London, UK with Dr. Ellen Solomon**

1989-1992

6/89 – 12/92 Postdoctoral Fellow

- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

**McMaster University
Hamilton, Ontario, Canada with Dr. G. D. Sweeney**

1983

5/83 – 8/83: NSERC Summer Student

- *In vitro* metabolism of β -naphthoflavone in C57BL/6J and DBA mice

EDUCATION

Ph.D.

"Phenotypic and genotypic effects of mutations in the human retinoblastoma gene."

Supervisor: Dr. R. A. Phillips

University of Toronto
Toronto, Ontario, Canada.
Department of Medical
Biophysics.

1989

Honours B.Sc

"The *in vitro* metabolism of the cytochrome P-448 inducer β -naphthoflavone in C57BL/6J mice."

Supervisor: Dr. G. D. Sweeney

McMaster University,
Hamilton, Ontario, Canada.
Department of Biochemistry

1983

ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76th Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

Goddard A, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

Goddard A, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

Goddard A, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

PUBLICATIONS

- Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD** Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.
- Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausek SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.
- Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.
- Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.
- Lee J. Ho WH. Maruoka M. Corpuz RT. Baldwin DT. Foster JS. **Goddard AD**. Yansura DG. Vandlen RL. Wood WI. Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.
- Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.
- Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.
- Guo S, Yamaguchi Y, Schilbach S, Wada T.; Lee J, **Goddard A**, French D , Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.
- Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.
- Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.
- Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.
- Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused; a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.
- Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

- Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.
- Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.
- Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* **59**: 2718-2723.
- Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.
- Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.
- Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.
- Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.
- Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.
- Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature.* **391**(6662): 90-92.
- Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* **7**(12): 1003-1006.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15-26.
- Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.

Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.

Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.

Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.

Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.

Goddard AD and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.

Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.

Goddard AD, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S-119S.

Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.

Foulkes W, **Goddard A.** and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.

Goddard AD, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.

Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.

Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the β -subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.

Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a NotI linking library from human chromosome 17q. *Genomics* **10**: 477-480.

Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* **46**: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* **62**: 394-408.

Goddard AD, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* **37**: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* **340**: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* **8**: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci. USA* **83**: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-myc expression. *Nature* **322**: 555-557.

Goddard AD, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* **152**: 31-38.

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

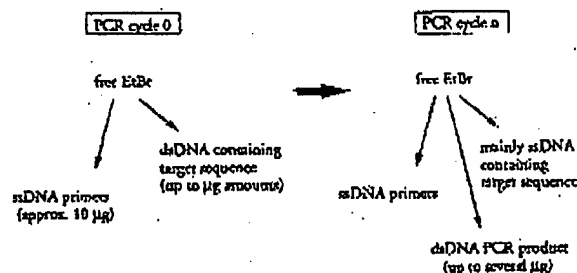


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

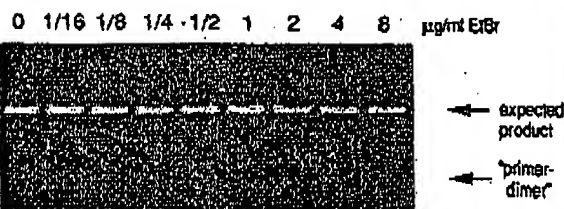


FIGURE 2 Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQ α , made in the presence of increasing amounts of EtBr (up to 8 μ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

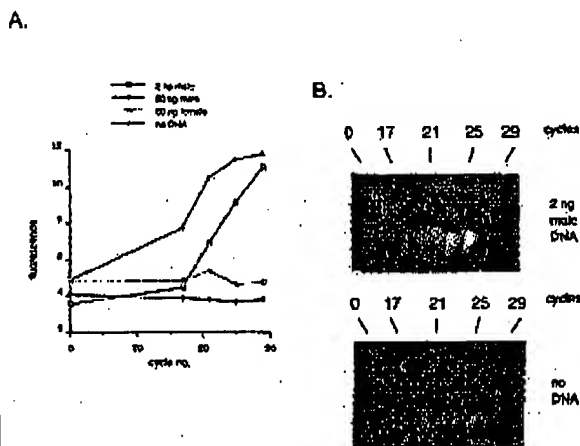


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photograph of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ss-DNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA DQ α gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 μ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β -globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β -globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β -globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β -globin individual (AA); from a heterozygous sickle β -globin individual (AS); and from a homozygous sickle β -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

smocy.

ess the
HLA
cent at
oncen-
lowing
e 2, gel
ie yield
Br was
ndicat.

fic se-
nent of
ries of
rimers
human
either
DNA.
fter 0,
r each
ts fluo-
plotted
of this
case in
DNA is
umber.
cc-fold
ontain-
ncrease
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
n a UV
h a red
ns that
VA.
-globin
sch has
etection
Figure
ications
graphy
These
for ci-
human
nparted
ual 3'
primer
has am-
c of the
ent.^{21,22}
nsists of
the (left
Three
zygous,
ozygous
ozygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

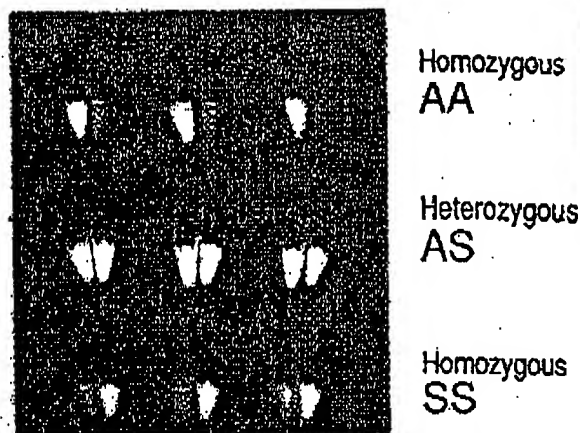


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCR's) for each input DNA.

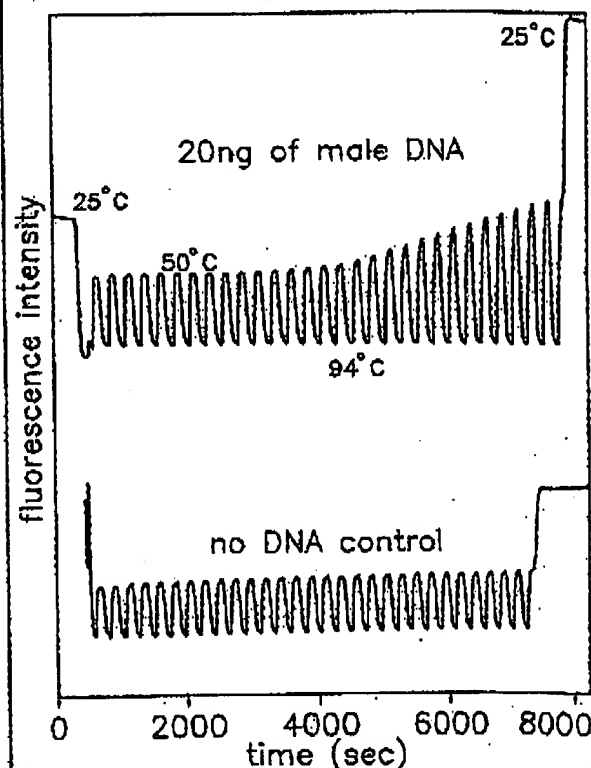


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube³, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiber optics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers (H26 and CH27)¹⁹ and approximately 10^5 copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁰, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/Hp14A (wild-type globin specific primers) or HGP2/Hp14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Crist, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQ* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1649-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Fieroz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Rien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Klugger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Helger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-38.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Hofford, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Froc, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Szczy, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3755-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:141-147.
- Abbott, M. A., Fieroz, B. J., Byrne, B. C., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. C., Doherty, M. and Giachier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:885-890.
- Wu, D. Y., Ugaz, L., Pal, B. R. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Fieroz, E., Williams, J. F. and Wondenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-109, 106-112.
- Tumosa, N. and Kahan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Med.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml detection limit: 1 ng/ml CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APRIL 1992

417

GENENTECH, INC.
1 DNA Way
South San Francisco, CA 94080 USA
Phone: (650) 225-1000

FAX: (650) 952-9881

FACSIMILE TRANSMITTAL

Date: 19 July 2004

To: Anna Barry
Heller Ehrman

Re: Higuchi reference

Fax No: 324-6638

From: Patty Tobin, Assistant to Elizabeth M. Barnes, Ph.D.
Genentech, Inc. Legal Department

Number of Pages including this cover sheet: 6

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger¹, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. ¹Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR¹ to clinical diagnostics are well known^{2,3}, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases⁴ made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization^{5,6}, gel electrophoresis with or without use of restriction digestion^{7,8}, HPLC⁹, or capillary electrophoresis¹⁰. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing¹¹.

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.¹², developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.¹³, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA¹⁴⁻¹⁶. As outlined in Figure 1, a prototypic PCR

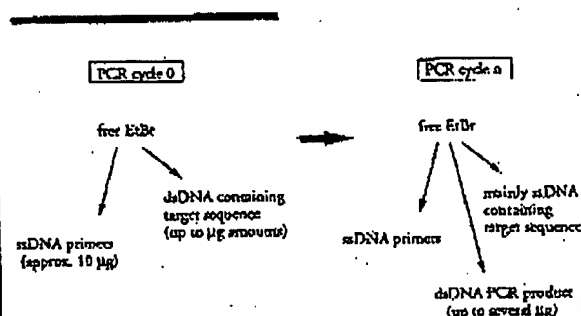


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

rotavirus fusion
Sci. USA 85:

drmann, T. A.,
immunology
to *Pseudomonas*

ngam, M. C.,
od of cloning
in *es* as single
1066-1070.

schberg, D. L.,
our effects of
isopentide phos-

s, G., Delcide,
penetia, A. A.
re-targeted by
1184-1188.
Vol. 2, p.

and Stevens,
with anti-viral
J. Biochem. J.

A. L., Carnicelli,
d properties of
oxidase activity.

ization of the
abix eukaryotic
528.

urification and
stacea americana
chem. Biophys.

1982. Purifica-
of the antiviral
bkwed). Bio-

L. Dodecandria,
dodecandria. Bio-

new inhibitor of
tein. 255:6947-

Abbondanza, A.,
ribosome-inacti-
(white bryony).

-synthesis inhib-
Lett. 159:209-

8. Isolation and
inhibitory pro-
chem. 52:1223-

v. L., Sperti, S.,
ctro by proteins
clon). Biochem.

nza, A., Cesini,
Purification and
rib RNA N-gly-
ation from the
Acta. 993:287-

Guillemot, J. C.,
1988. Trichold-
of *Trichostema*
8.

itors of animal
Biophys. Acta

roperties of the
ozin inhibiting

of abrin: a toxic
ferent biological
ur. J. Biochem.

Franz, H. 1980.
i *Vicia alba* L.

i. and Sirpe, F.
d madoecia, the

., Brown, A. N.,
s of volkmain, a
14589-14995.
nd properties of
nistry 18:2615-

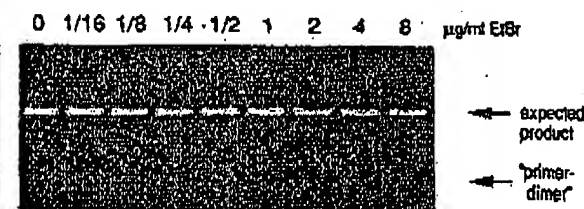


FIGURE 2 Gel electrophoresis of PCR amplification products of the human, nuclear gene, HLA DQ α , made in the presence of increasing amounts of EtBr (up to 8 μ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

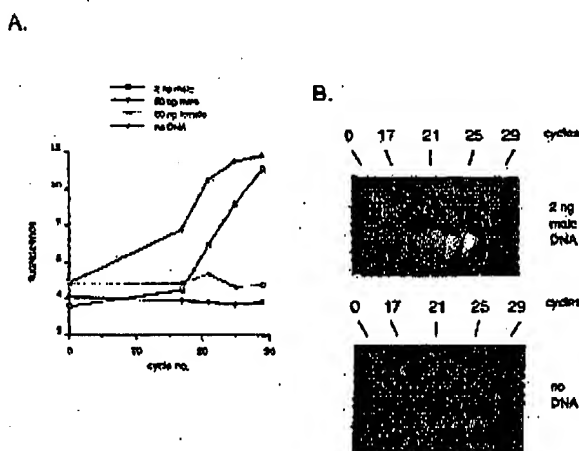


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA¹⁷ to micrograms per PCR¹⁸. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA DQ α gene¹⁹ were performed with the dye present at concentrations from 0.06 to 8.0 μ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome²⁰. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β -globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β -globin gene²¹. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β -globin allele present^{21,22}.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous wild-type β -globin individual (AA); from a heterozygous sickle β -globin individual (AS); and from a homozygous sickle β -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

monoc.

ess the
HLA
cent at
oncen-
lowing
e 2, gel
ic yield
Br was
ndicat.

fic se-
nent of
rics of
rimers
human
either
DNA.
fter 0,
or each
is fluo-
plotted
of this
case in
DNA is
umber.
ce-fold
ontain-
increase
her no
DNA
fewer
in fluo-
f these
the ex-
taining
in the

ualized
n a UV
b a red
ns that
VA.
-globin
sch has
etection
Figure
ications
graphy
These
for ci-
human
nparted
ual 3'
primer
hus am-
e of the
ent^{21,22}
nsists of
the (left
, Three
zygous,
zygous
zygous
genomic
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells⁶. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

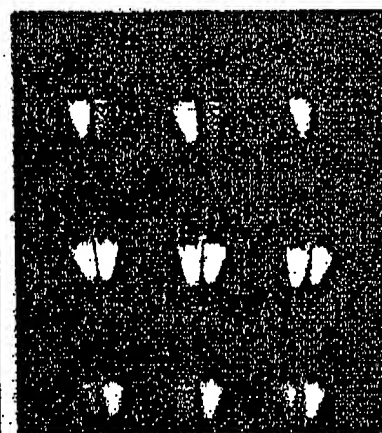


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

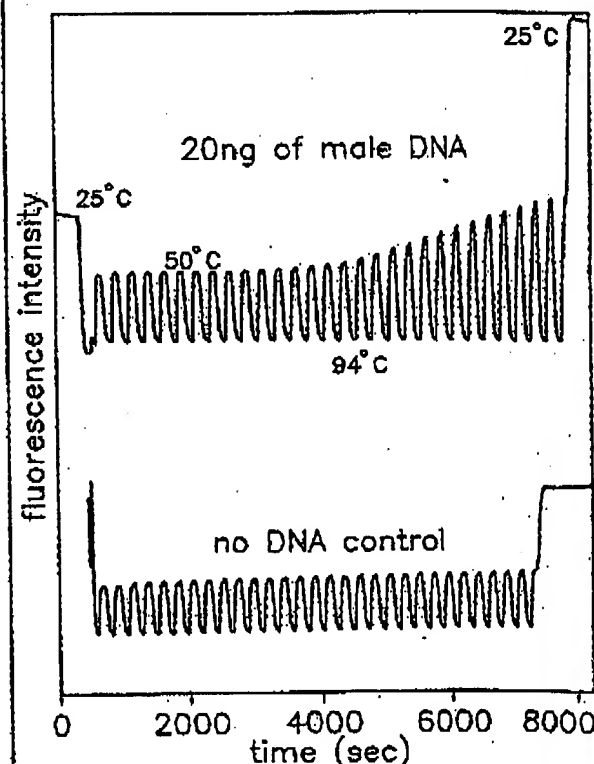


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the “primer-dimer” artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube³, and the “hot-start”, in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins²³. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^3 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5′ “add-on” to the oligonucleotide primer²⁴.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format²⁵. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader²⁶.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQ α gene amplifications containing EtBr. PCRs were set up in 100 μ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl₂; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ α gene specific oligonucleotide primers (H26 and CH27)¹⁹ and approximately 10^3 copies of DQ α PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a “step-cycle” program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

Y-chromosome specific PCR. PCRs (100 μ l total reaction volume) containing 0.5 μ g/ml EtBr were prepared as described for HLA-DQ α , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2²⁰, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a “step-cycle” program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β -globin gene PCR. Amplifications of 100 μ l volume using 0.5 μ g/ml of EtBr were prepared as described for HLA-DQ α above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.²¹. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a “step-cycle” program. An annealing temperature of 55°C had been shown by Wu et al.²¹ to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 25A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Crist, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with “5 minute-epoxy” to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQ* 51:269-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Celfand, D. and Smitsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Celfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6240-6244.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:293-294.
- Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-548.
- Hedger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Szczy, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:3519-3534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ¹H NMR spectroscopy. *Nuc. Acids Res.* 18:3753-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Smitsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. C., Doherty, M. and Gitshier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Uguzoz, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spack, D., Guda, L., Levenson, C. and Smitsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Submitted*.
- Higuchi, R. 1989. Using PCR to engineer DNA, p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Fozzard, E., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-109, 106-112.
- Tumosa, N. and Kahwa, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

IBL

IMMUNO BIOLOGICAL LABORATORIES

sCD-14 ELISA

Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations.

During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml detection limit: 1 ng/ml CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/49100 61-64 · FAX +40/40 11 98

BIOTECHNOLOGY VOL 10 APRIL 1992

417

Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5'→3' nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3'-end nucleotide. In all cases, the reporter dye was attached to the 5' end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3'-end nucleotide exhibited a larger signal in the 5' nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3'-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.⁽¹⁾ The assay exploits the 5'→3' nucleolytic activity of *Taq* DNA polymerase^(2,3) and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5' end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).^(4,5) During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5'→3' nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5' nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.⁽⁶⁾ Because of this, it has been assumed that the quencher dye must be attached near the 5' end. Surprisingly, we have found that attaching a rhodamine dye at the 3' end of a probe still provides adequate quenching for the probe to perform in the 5' nuclease

PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5' end and a quencher dye on the 3' end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

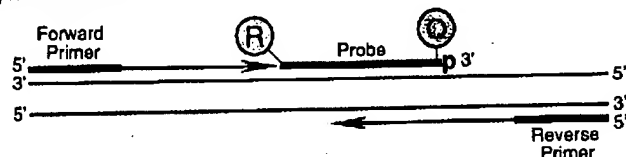
MATERIALS AND METHODS

Oligonucleotides

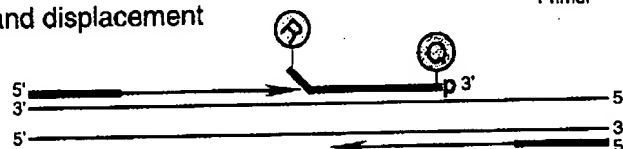
Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3'-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5' end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3' end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

Research

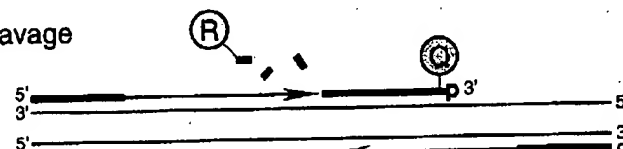
Polymerization



Strand displacement



Cleavage



Polymerization completed

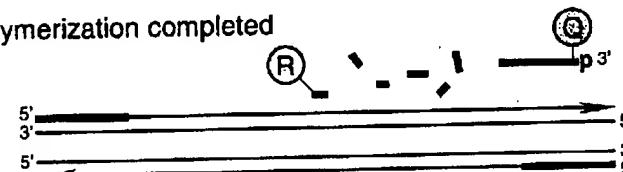


FIGURE 1 Diagram of 5' nuclease assay. Stepwise representation of the 5' → 3' nucleolytic activity of Taq DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mm Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C₈ 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5' end.

PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil N-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin

gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.)⁽⁷⁾ was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.⁽⁸⁾ Actin amplification reactions contained 4 mM MgCl₂, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each

primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM MgCl₂, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
F119	primer	ACCCACAGGAAGTATCACCACCTC
R119	primer	ATGTCGCGTTCGGCTGACGTTCTGC
P2	probe	TCGCATTACTGATCGTIGCCAACCACTp
P2C	complement	GTACTGCTTGGCAACGATCAGTAATGCGATG
P5	probe	CGGAITTCCTGGTATCTATGACAAGGAtp
P5C	complement	TTCATCCTTGTCTATAGATACCAGCAATCCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCGGAACCGCTCATTGCCAATGG
A1	probe	ATGCCCTCCCCCATGCCATCCTTCGCTp
A1C	complement	AGACGCAGGATGGCATGGGGGAGGGCATAc
A3	probe	CGCCCTGGACTTCGAGCAAGAGATp
A3C	complement	CCATCTCTTGCTCGAAGTCCAGGGCGAC

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5' → 3' direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5' nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3' phosphate on each probe.

A1-2 RAQGCCCTCCCCATGCCATCCTGCGTp
 A1-7 RATGCCCTCCCCATGCCATCCTGCGTp
 A1-14 RATGCCCTCCCCATGCCATCCTGCGTp
 A1-19 RATGCCCTCCCCATGCCATCCTGCGTp
 A1-22 RATGCCCTCCCCATGCCATCCTGCGTp
 A1-26 RATGCCCTCCCCATGCCATCCTGCGTp

Probe	518 nm		582 nm		RQ ⁻	RQ ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	25.5 ± 2.1	32.7 ± 1.9	38.2 ± 3.0	38.2 ± 2.0	0.67 ± 0.01	0.86 ± 0.06	0.19 ± 0.06
A1-7	53.5 ± 4.3	395.1 ± 21.4	108.5 ± 6.3	110.3 ± 5.3	0.49 ± 0.03	3.58 ± 0.17	3.09 ± 0.18
A1-14	127.0 ± 4.9	403.5 ± 19.1	109.7 ± 5.3	93.1 ± 6.3	1.16 ± 0.02	4.34 ± 0.15	3.18 ± 0.15
A1-19	187.5 ± 17.9	422.7 ± 7.7	70.3 ± 7.4	73.0 ± 2.8	2.67 ± 0.05	5.80 ± 0.15	3.13 ± 0.16
A1-22	224.6 ± 9.4	482.2 ± 43.6	100.0 ± 4.0	96.2 ± 9.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 8.9	454.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.08	3.29 ± 0.08

FIGURE 2 Results of 5' nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 s.d. for six reactions run without added template (no temp.) and six reactions run with template (+ temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ⁻ and RQ⁺ values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ⁻) from the RQ value for the complete reaction including template (RQ⁺).

RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target

sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for Taq polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are

clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ⁻ depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ⁻ values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ⁻ values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

TABLE 2 Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

Probe	518 nm		582 nm		RQ ⁻	RQ ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	116.2 ± 6.4	115.6 ± 2.5	0.47 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	82.8 ± 4.4	384.0 ± 34.1	105.1 ± 6.4	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	555.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.88 ± 0.10
P5-10	77.5 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.05	2.55 ± 0.06	1.66 ± 0.08
P5-28	64.0 ± 5.2	333.6 ± 12.1	100.6 ± 6.1	94.7 ± 6.3	0.63 ± 0.02	3.53 ± 0.12	2.89 ± 0.13

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to Fig. 2.

Research

fluorescence of a reporter dye on the 5' end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5' nuclease PCR assay.

To test the hypothesis that quenching by a 3' TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5' end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3' end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5' and 3' ends from being in proximity.

When TAMRA is placed toward the 3' end, there is a marked Mg^{2+} effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of Mg^{2+} concentration. With TAMRA attached near the 5' end (probe A1-2 or A1-7), the RQ value at 0 mM Mg^{2+} is only slightly higher than RQ at 10 mM Mg^{2+} . For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM Mg^{2+} are very high, indicating a much

reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM Mg^{2+} followed by a gradual decline as the Mg^{2+} concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM Mg^{2+} with a gradual decline at higher Mg^{2+} concentrations. In a low-salt environment with no Mg^{2+} present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of Mg^{2+} ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3' end is close to the 5' end. Therefore, the observed Mg^{2+} effects support the notion that quenching of a 5' reporter dye by TAMRA at or near the 3' end depends on the flexibility of the oligonucleotide.

DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5' end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5' end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what

matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3' end, or any other position, can quench 6-FAM at the 5' end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5' nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5' nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5' nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of RQ^- , which is the ratio of reporter to quencher fluorescent emissions for a no template control PCR. Influences on the value of RQ^- include the particular reporter and quencher

TABLE 3 Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

Probe	518 nm		582 nm		RQ	
	ss	ds	ss	ds	ss	ds
A1-7	27.75	68.53	61.08	138.18	0.45	0.50
A1-26	43.31	509.38	53.50	93.86	0.81	5.43
A3-6	16.75	62.88	39.33	165.57	0.43	0.38
A3-24	30.05	578.64	67.72	140.25	0.45	3.21
P2-7	35.02	70.13	54.63	121.09	0.64	0.58
P2-27	39.89	320.47	65.10	61.13	0.61	5.25
P5-10	27.34	144.85	61.95	165.54	0.44	0.87
P5-28	33.65	462.29	72.39	104.61	0.46	4.43

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM $MgCl_2$. (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of $MgCl_2$, 120 μ l of each sample was heated at 95°C for 5 min. Following the addition of 80 μ l of 25 mM $MgCl_2$, each sample was allowed to cool to room temperature and the fluorescence emissions were measured. Reported values are the average of three determinations.

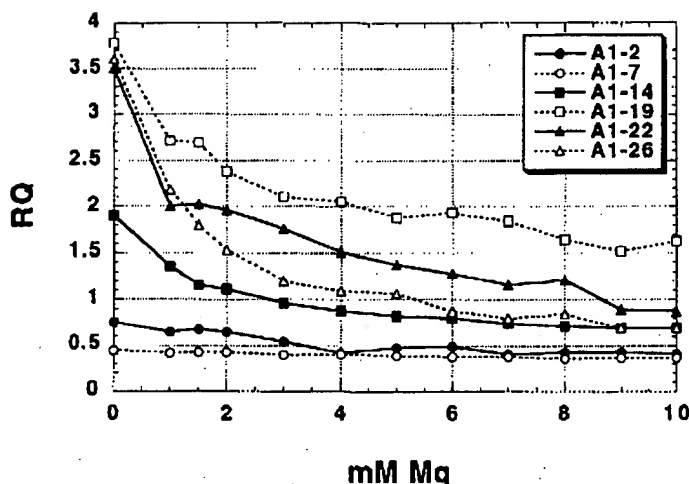


FIGURE 3 Effect of Mg^{2+} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (upper right) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m , presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.⁽¹⁾

The rise in RQ^- values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed quencher. For the other three sets of

probes, the interpretation of RQ^- values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ^- value. For the P5 probes, the RQ^- for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ^- .

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3' end. This illus-

trates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.⁽⁹⁾ This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.⁽¹¹⁾ demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the $\Delta F508$ mutant. Their probes had TAMRA attached to the seventh nucleotide from

Research

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.⁽¹⁰⁾ describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**: 3761-3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction prod-

uct by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* **88**: 7276-7280.

3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**: 778-783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**: 55-75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303-339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* **58**: 719-726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Inter-species homology of sequences in the introns. *Proc. Natl. Acad. Sci.* **82**: 6133-6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of β -actin-specific mRNA transcripts using xeno-competitive PCR. *PCR Methods Applic.* **3**: 57-59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* **22**: 2424-2425.

Received December 20, 1994; accepted in revised form March 6, 1995.

THIS MATERIAL MAY BE PROTECTED
BY COPYRIGHT LAW (17 U.S. CODE)

GENOME METHODS

Real Time Quantitative PCR

Christian A. Heid,¹ Junko Stevens,² Kenneth J. Livak,² and
P. Mickey Williams^{1,3}

¹BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080; ²Genentech, Inc., 460 Point San Bruno Blvd., South San Francisco, California 94080

²Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genomic quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Platak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)-PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative

that it be used properly for quantitation (Rasmackers 1995). Many early reports of quantitative PCR and RT-PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Perre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT-PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as β -actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)-PCR, has been developed and is used widely for PCR quantitation. QC-PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Ptacek et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

³Corresponding author.

REAL TIME QUANTITATIVE PCR

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QPCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labelling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fusco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5' nuclease assay first described by Holland et al. (1991). The method uses the 5' nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Livak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI-Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

RESULTS

PCR Product Detection in Real Time

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pF8TM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human β -actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed ΔRn (or $\Delta R(2)$) using the following equation: $\Delta Rn = (Rn^1) / (Rn^2)$, where Rn^1 = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and Rn^2 = emission intensity of re-

HUI ET AL.

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points (ΔRn s) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The ΔRn mean value is plotted on the y-axis, and time, represented by cycle number, is plotted on the x-axis. During the early cycles of the PCR amplification, the ΔRn

value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which

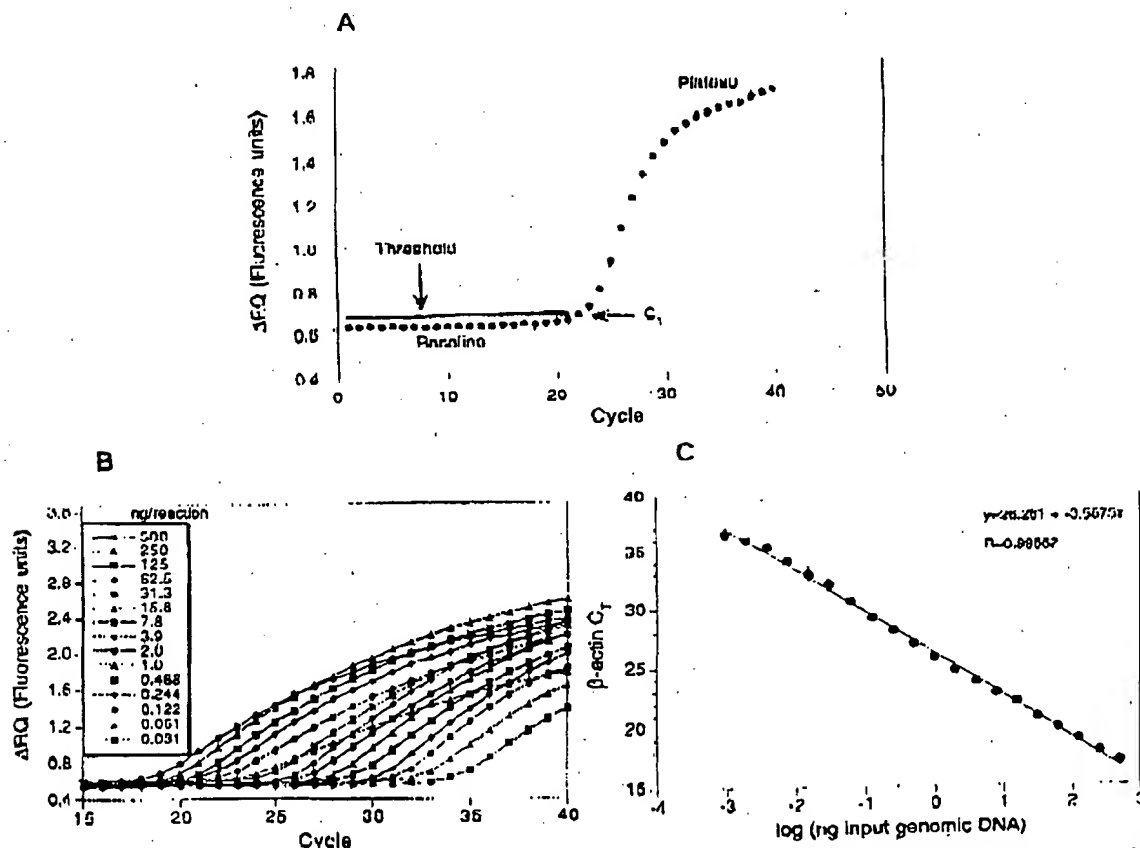


Figure 1 PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. C_t values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with β -actin primers. (C) Input DNA concentration of the samples plotted versus C_t . All

REAL TIME QUANTITATIVE PCR

the amplification plot crosses the threshold is defined as C_T . C_T is reported as the cycle number at this point. As will be demonstrated, the C_T value is predictive of the quantity of input target.

C_T Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human β actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the C_T values. Figure 1C represents the C_T values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The C_T values decrease linearly with increasing target quantity. Thus, C_T values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated C_T value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar C_T values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using C_T values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se-

ments over a very large range of relative starting target quantities.

Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing β -actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of C_T values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean C_T values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for β -actin gene quantity. The highest C_T difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured β -actin C_T value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected C_T value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

Quantitative Analysis of a Plasmid After

7000 001 888 VVJ AC:BT 7007/00/7T

III) FINAL

Table 1. Reproducibility of Sample Preparation Method

Sample no.	100 ng				25 ng			
	C _T	mean	standard deviation	CV	C _T	mean	standard deviation	CV
1	18.24 18.23 18.33	18.27	0.06	0.32	20.48 20.55 20.5	20.51	0.03	0.17
2	18.33 18.35 18.44				20.61 20.59 20.41			
3	18.3 18.3 18.42	18.37	0.06	0.32	20.54 20.6 20.49	20.54	0.11	0.54
4	18.15 18.23 18.32				20.48 20.44 20.38			
5	18.4 18.38 18.46	18.34	0.07	0.36	20.68 20.87 20.63	20.43	0.05	0.26
6	18.54 18.67 19				21.09 21.04 21.01	20.73	0.13	0.61
7	18.28 18.36 18.52	18.71	0.24	1.26	20.67 20.73 20.65			
8	18.45 18.7 18.73				20.98 20.84 20.75	20.68	0.04	0.2
9	18.18 18.34 18.36	18.63	0.16	0.83	20.46 20.54 20.48			
10	18.42 18.57 18.66				20.79 20.78 20.62	20.86	0.12	0.57
		18.29	0.1	0.55				
						20.51	0.07	0.32
Mean	(1 10)	18.12	0.17	0.90		20.73	0.1	0.16
						20.66	0.19	0.94

tor containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 µg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β -Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β -actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β -actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β -actin C_T values of the triplicates were plotted (error bars represent standard deviation). The highest difference

between any two sample means was 0.95 C_T. Ten nanograms of total DNA of each sample were also examined for β -actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β -actin C_T value difference was 1.0. As Figure 3 shows, the rate of β -actin C_T change between the 100 and 10-ng samples was similar (slope values range between -3.56 and -3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

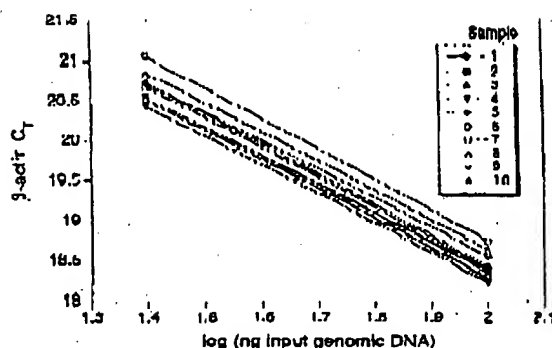


Figure 2 Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs. C_T . In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean β -actin C_T value obtained for each 100-ng sample on a β -actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample, a , was obtained by extrapolation to the x-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate

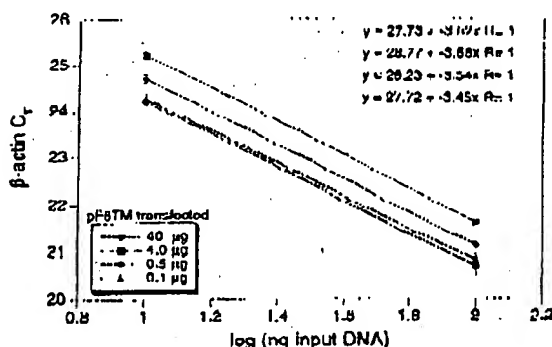


Figure 3 Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1 μ g of pF8TM) were analyzed for the β -actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the β -actin C_T values are plotted versus the total input DNA concentration.

REAL TIME-QUANTITATIVE PCR

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean C_T values increase) with decreasing amounts of plasmid (transfected). The mean C_T values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM, b , found in each of the four transfections was determined by extrapolation to the x-axis of the standard curve in Figure 4B. These uncorrected values, b , for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100 \text{ ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

where a = actual genomic DNA in a sample and b = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40 μ g of plasmid, was 35 μ g per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e., β -actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

HUDD L1 AL.

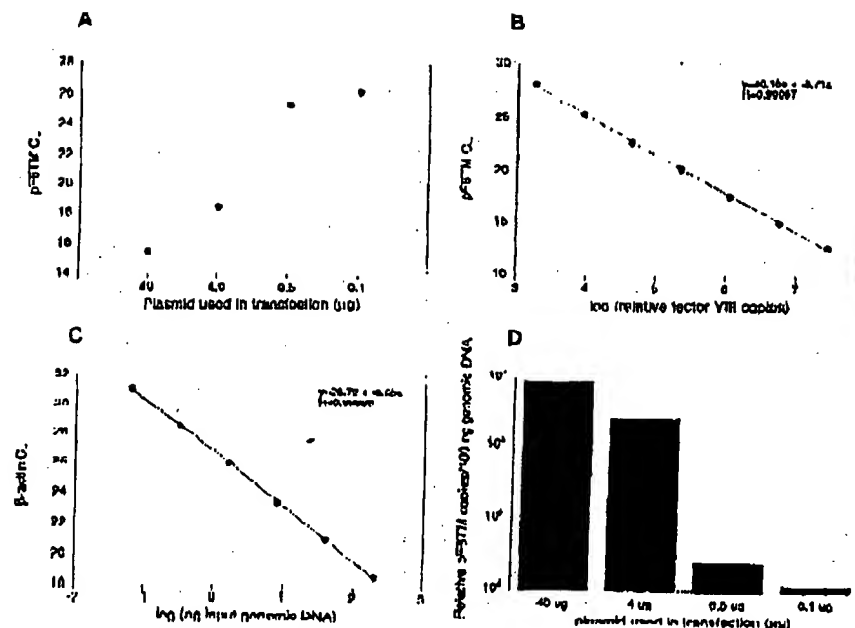


Figure 4 Quantitative analysis of pF8TM in transfected cells. (A) Amount of plasmid DNA used for the transfection plotted against the mean C_T value determined for pF8TM remaining 24 hr after transfection. (B,C) Standard curves of pF8TM and β -actin, respectively. pF8TM DNA (B) and genomic DNA (C) were diluted serially 1:5 before amplification with the appropriate primers. The β -actin standard curve was used to normalize the results of A to 100 ng of genomic DNA. (D) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β -actin) for quantitative PCR or house-keeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multi-gene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed

for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C_T , correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications [quantitative gene expression (RT-PCR), gene copy assays (Her2, HIV, etc.), genotyping (knockout mouse analysis), and immunoprecipitation].

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes--greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

METHODS

Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c251 (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR (GeneAmp 12. (1)th RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)) using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *HindIII* and *HindIII* restriction site sequences at the 5' end) and cloned into pCDM-32 (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-CCCCCTGKCAAGAGCTGATCTGTC-3' and F8rev 5'-AAACCTTCAACCTGGATGCTAGCT-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 3' untranslated region of the parent pCIS2.8c251 plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human β -actin gene was amplified with the primers β -actin forward primer 5'-TCACCCACACTCTT GCCCATCTACCA-3' and β -actin reverse primer 5'-CAGCCGAACCGCTTCATGCGCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50 μ l) contained a DNA sample, 10 \times PCR Buffer II (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dTTP, 4 mM MgCl₂, 1.25 Units Ampil Tag DNA polymerase, 0.5 unit AmpErase uracil N-glycosylase (UNG), 60 pmole of each factor VIII primer, and 15 pmole of each β -actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'-(FAM)AGCTCTTCACACCTGCTCTCTTCTCTTGCCTT(TAMRA)p 3' and β -actin probe 5'-(FAM)ATGCCCTX(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801 0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector Instrument manual. Briefly, probe T_m should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with

REAL TIME QUANTITATIVE PCR

reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S F12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1 μ g were added to 1.5 ml of a solution containing 0.125 M CaCl₂ and 1 \times HEPES. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO₂ for 24 hr, washed with PBS, and resuspended in PBS. The resuspended cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200 μ l of 30 mM Tris-HCl at pH 8.0.

ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bassler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Koon, and C.A. Ball. 1995. Use of a fluorescent probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *App. Environ. Microbiol.* 61: 3724-3729.
- Buckner-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189-201.
- Clement, M., S. Menzo, P. Baguette, A. Manzin, A. Valenza, and P.E. Varaldo. 1993. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 197-196.
- Connor, R.J., H. Mohut, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4⁺ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772-1777.
- Eaton, D.L., W.J. Wood, D. Eaton, P.H. Nass, P.

HFID LI AL

- Venar, and C. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343-8347.
- Rasco, M.J., C.P. Treanor, S. Spivack, H.L. Wigge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140-147.
- Perre, B. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1-9.
- Fortado, M.R., L.A. Kingsley, and S.M. Wollinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2092-2100.
- Gibson, D.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).
- Gorman, C.M., D.R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3-10.
- Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Molecular Biology* 10: 413-417.
- Holland, P.M., R.D. Abramson, R. Watson, and D.J. Gelfand. 1991. Detection of specific polymerase chain reaction products by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276-7280.
- Huang, S.K., H.Q. Xiao, T.J. Klein, G. Paciotti, H.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688-2694.
- Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen. *Am J. Immun.* 154: 6157-6162.
- Kellogg, D.E., J.J. Shinsky, and S. Kwak. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202-208.
- Lee, J.-G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761-3766.
- Livak, K.J., S.J. Flood, J. Marmaro, W. Chu, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357-362.
- Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards fully automated genome-wide polymorphism screening. [Letter] *Nature Genet.* 9: 341-342.
- Mulder, J., N. McKinney, C. Christopherson, J. Shinsky, L. Greenfield, and S. Kwak. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292-300.
- Pang, S., Y. Koyanagi, S. Mills, C. Wiloy, H.V. Vinters, and L.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85-89.
- Platak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1995a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *HiTechniques* 14: 70-81.
- Platak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hahn, G.M. Shaw, and J.D. Lifson. 1995b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749-1754.
- Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495-503.
- Racymackers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91-94.
- Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750-768.
- Slamon, T.J., G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, and W.J. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science* 235: 177-182.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. PAI and TNF increase the precursor of Nk-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157-162.
- Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201-5205.
- Williams, S., C. Schwer, A. Krishnarao, C. Heid, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA^{*†}, TODD A. SWANSON^{*}, JAMES W. WELSH^{*}, MARGARET A. ROY[‡], DAVID A. LAWRENCE^{*}, JAMES LEE[‡], JENNIFER BRUSH[‡], LISA A. TANEYHILL[§], BETHANNE DEUEL[‡], MICHAEL LEW[¶], COLIN WATANABE[¶], ROBERT L. COHEN^{*}, MONA F. MELHEM^{**}, GENE G. FINLEY^{**}, PHIL QUIRKE^{††}, AUDREY D. GODDARD[‡], KENNETH J. HILLAN[¶], AUSTIN L. GURNEY[‡], DAVID BOTSTEIN^{‡,‡‡}, AND ARNOLD J. LEVINE[§]

Departments of ^{*}Molecular Oncology, [‡]Molecular Biology, [§]Scientific Computing, and [¶]Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; ^{**}University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ^{††}University of Leeds, Leeds, LS29JT United Kingdom; ^{‡‡}Department of Genetics, Stanford University, Palo Alto, CA 94305; and [§]Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

[†]To whom reprint requests should be addressed. e-mail: diane@gene.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0
PNAS is available online at www.pnas.org.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{\Delta\Delta Ct}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The Δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of $\approx 40,000$ (M_r , 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of $\approx 27,000$ (M_r , 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

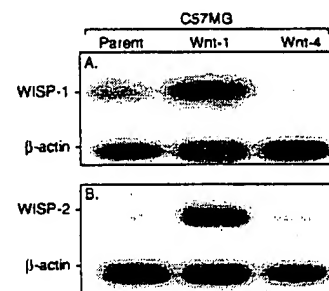


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

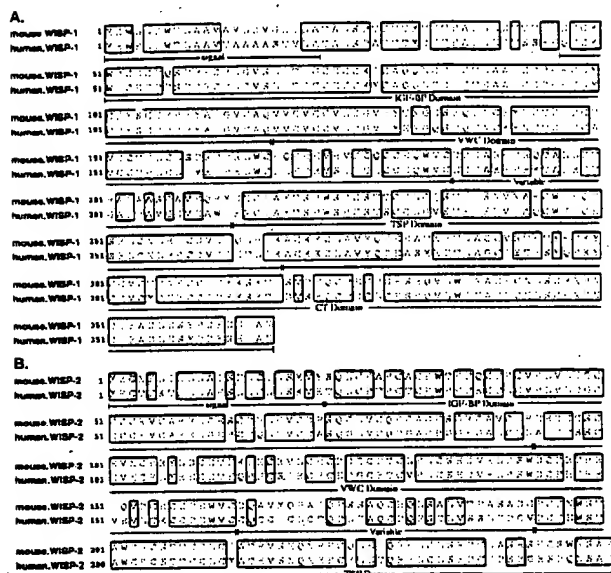


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

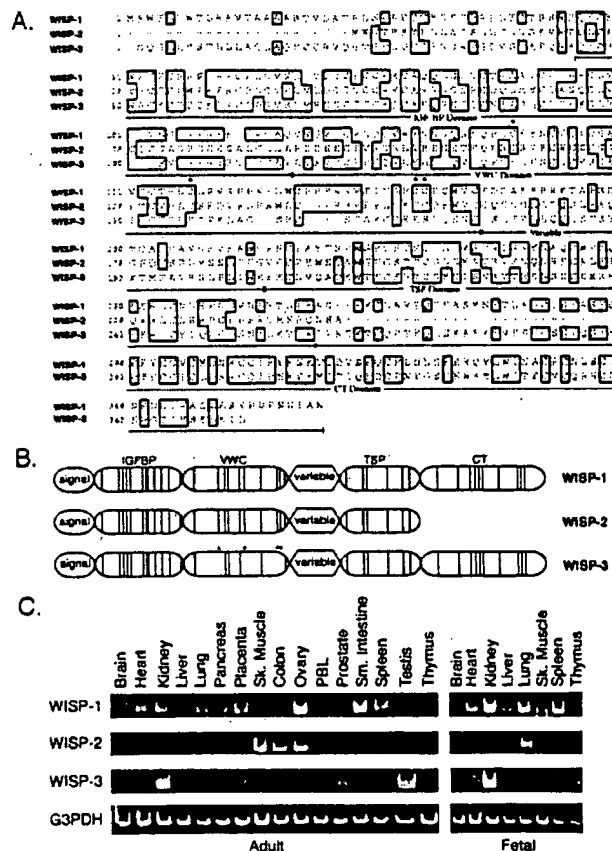


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCsxC motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

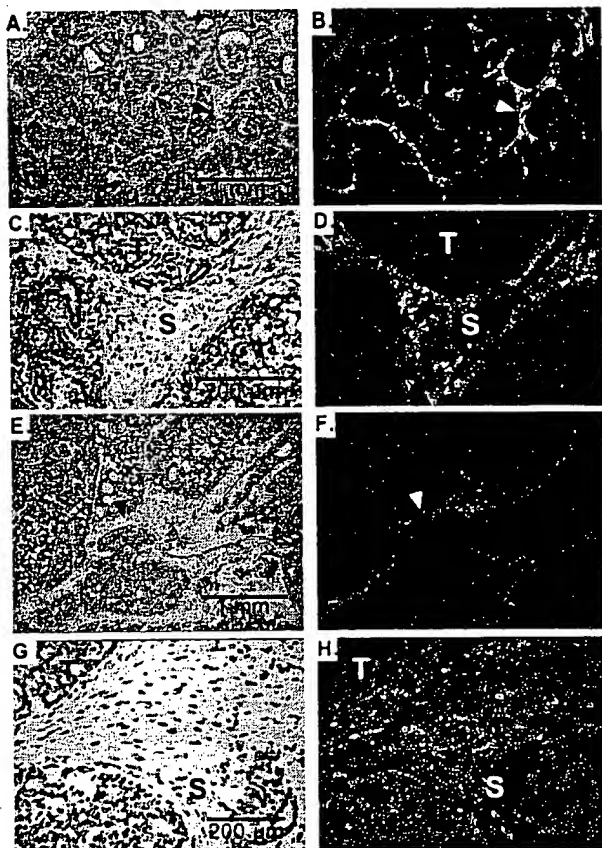


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene MYB (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

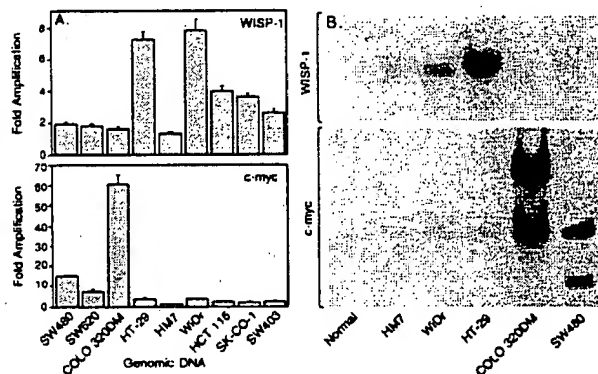


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

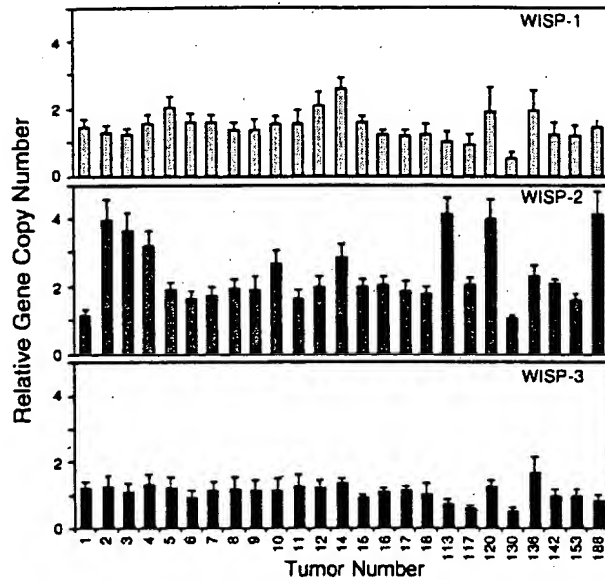


Fig. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

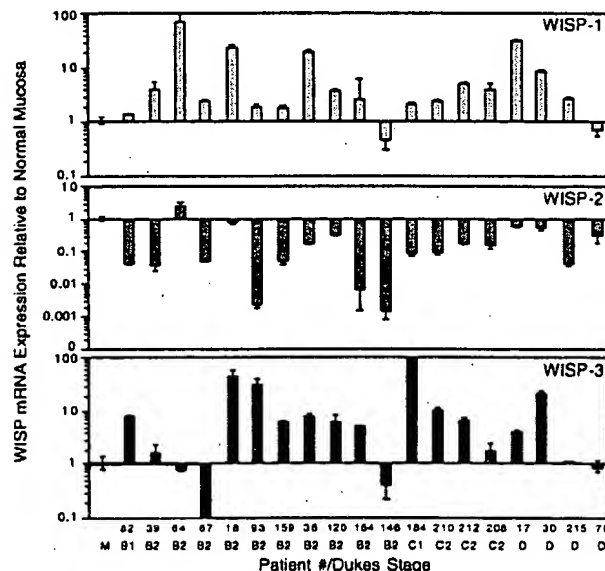


Fig. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_3\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of WISP-1 and WISP-2 in the stromal cells of breast tumors supports this paracrine model.

An analysis of WISP-1 gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human WISP-2 was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for WISP-2 may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of WISP-2, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which WISP-2 RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of WISP-2 in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the WISP genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the WISPs. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of WISPs as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the WISPs in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125–130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* 57, 109–111.
- Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435–449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* 54, 4257–4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* 17, 3052–3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described^{20,21}. Briefly, after antigen pulsing (30 µg ml⁻¹ TTCF) with tetrapeptides (1–2 mg ml⁻¹), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with 1 µCi of ³H-thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating 200 µg TTCF with 0.25 µg pig kidney legumain in 500 µl 50 mM citrate buffer, pH 5.5, for 1 h at 37°C. **Glycopeptide digestions.** The peptides HIDNEED1, HIDN(N-glucosamine) EED1 and HIDNESD1, which are based on the TTCF sequence, and QQQHFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQHFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography¹¹. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30°C with 5–50 mU ml⁻¹ pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml⁻¹ α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. et al. Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098 (1997).
- Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208–213 (1993).
- Dalton, J. P., Hla Jmriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575–580 (1995).
- Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519–1524 (1992).
- Riese, R. J. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366 (1996).
- Rodriguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894–2898 (1992).
- Hewitt, E. W. et al. Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693–4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821–850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93–102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377–382 (1997).
- Lu, J. & van Halbeek, H. Complete ¹H and ¹³C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1–21 (1996).
- Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50–54 (1996).
- Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295–298 (1997).
- Wyatt, R. et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705–711 (1998).
- Botarelli, P. et al. N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128–3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101–4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535–559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043–1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537–539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti^{††}, Scot A. Marsters^{††}, David A. Lawrence^{††}, Margaret Roy^{*}, Frank C. Kischkel^{*}, Patrick Dowd^{*}, Arthur Huang^{*}, Christopher J. Donahue^{*}, Steven W. Sherwood^{*}, Daryl T. Baldwin^{*}, Paul J. Godowski^{*}, William I. Wood^{*}, Austin L. Gurney^{*}, Kenneth J. Hillan^{*}, Robert L. Cohen^{*}, Audrey D. Goddard^{*}, David Botstein[‡] & Avi Ashkenazi^{*}

^{*} Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

[‡] Department of Genetics, Stanford University, Stanford, California 94305, USA

^{††} These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells¹. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (Dcr3), that binds to FasL and inhibits FasL-induced apoptosis. The Dcr3 gene was amplified in about half of 35 primary lung and colon tumours studied, and Dcr3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily². Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (Dcr3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)³, Dcr3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of Dcr3 in mammalian cells; Dcr3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). Dcr3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of Dcr3 and OPG are conserved; however, the carboxy-terminal portion of Dcr3 is 101 residues shorter.

We analysed expression of Dcr3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high Dcr3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of Dcr3, we generated a recombinant, Fc-tagged Dcr3 protein. We tested binding of Dcr3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). Dcr3-Fc showed a significant increase in binding to cells transfected with FasL⁴ (Fig. 2a), but not to cells transfected with TNF⁵, Apo2L/TRAIL^{6,7}, Apo3L/TWEAK^{8,9}, or OPGL/TRACE/

RANKL¹⁰⁻¹² (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at $\sim 0.1 \mu\text{g ml}^{-1}$. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process¹. Consistent with previous results¹³, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes¹⁴⁻¹⁶. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from $\sim 65\%$ to $\sim 30\%$, with half-maximal inhibition at $\sim 1 \mu\text{g ml}^{-1}$; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL¹⁷.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

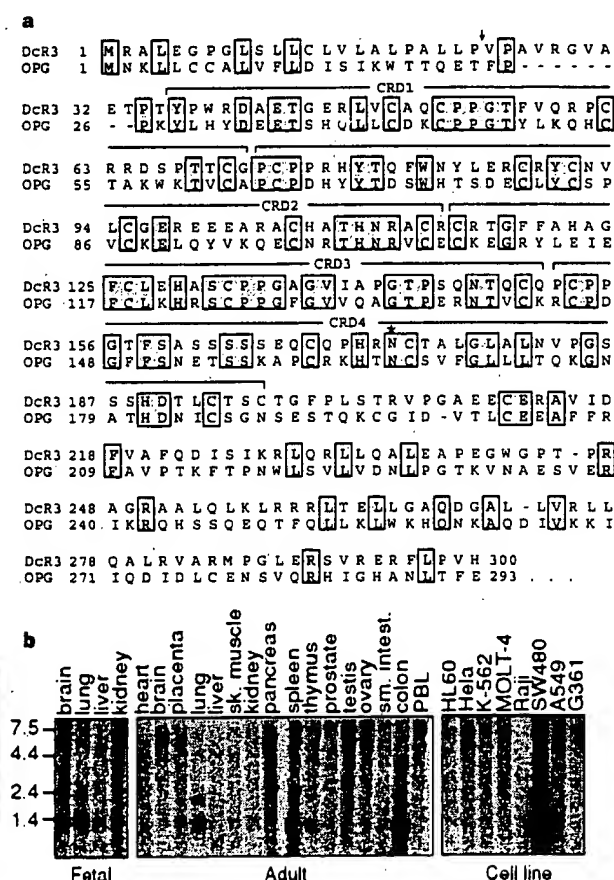


Figure 1 Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)⁺ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

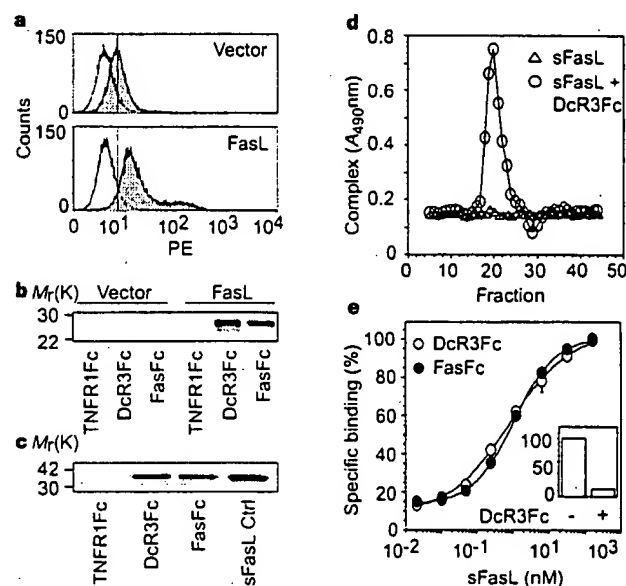


Figure 2 Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)¹⁸ in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epi-centre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG^{2,19}.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas²⁰. A second mechanism involves proteolytic shedding of FasL from the cell surface¹⁷. DcR3 competes with Fas for

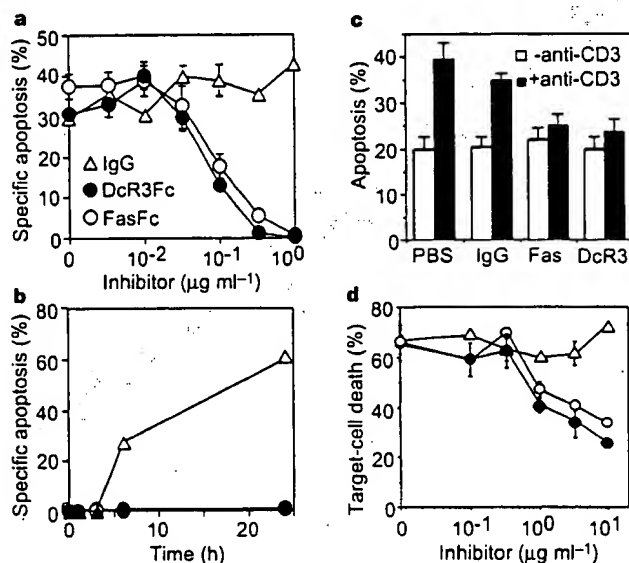


Figure 3 Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml⁻¹) oligomerized with anti-Flag antibody (0.1 μg ml⁻¹) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml⁻¹ DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml⁻¹). After 16 h, apoptosis of CD4⁺ cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with ⁵¹Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ⁵¹Cr (mean ± s.d. for two donors, each in triplicate).

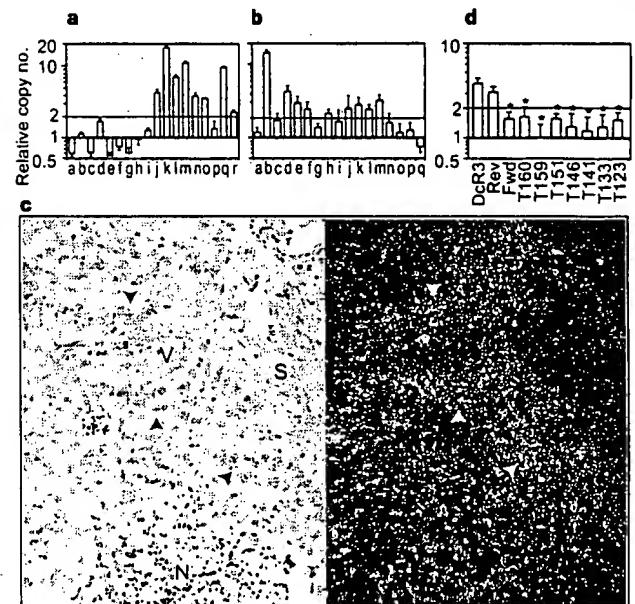


Figure 4 Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (**b**). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described²¹. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L²². Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG³, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L¹⁹. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response². Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in LifeseqTM (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described²³.

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL⁴ (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3-Fc or TNFR1-Fc, and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov-Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3-Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [³⁵S]cysteine and [³⁵S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3-Fc, Fas-Fc or TNFR1-Fc (5 µg), followed by protein A-Sepharose (Repligen). The precipitates were resolved by SDS-PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A-Sepharose, resolved by SDS-PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3-Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3-Fc-FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells pre-coated with anti-human IgG (Boehringer) to capture DcR3-Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin-horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3-Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3-Fc or Fas-Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas-Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3-Fc.

T-cell AICD. CD3⁺ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotec), stimulated with phytohaemagglutinin (PHA; 2 µg ml⁻¹) for 24 h, and cultured in the presence of interleukin-2 (100 U ml⁻¹) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4⁺ cells²⁴.

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotec), and incubated for 16 h with ⁵¹Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3-Fc, Fas-Fc or human IgG1. Target-cell death was determined by release of ⁵¹Cr in effector-target cocultures relative to release of ⁵¹Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR¹⁸ using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218x7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5'-CTTCTTCGCGCAGCTG-3' and 5'-ATCACGCCGGCACCAG-3' and the fluorogenic probe sequence was 5'-(FAM-ACACGATGCGTGCTCCAGCAG AAp-(TAMARA), where FAM is 5'-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2^(ΔCT), where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

- Nagata, S. Apoptosis by death factor. *Cell* 88, 355-365 (1997).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959-962 (1994).
- Simonet, W. S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309-319 (1997).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169-1178 (1993).
- Pennica, D. et al. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724-729 (1984).
- Pitti, R. M. et al. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687-12690 (1996).
- Wiley, S. R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673-682 (1995).
- Marsters, S. A. et al. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
- Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401-32410 (1997).
- Wong, B. R. et al. TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190-25194 (1997).
- Anderson, D. M. et al. A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175-179 (1997).
- Lacey, D. L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165-176 (1998).
- Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/Fas/CD95. *Nature* 373, 438-441 (1995).
- Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235-1238 (1995).
- Medvedev, A. E. et al. Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394-404 (1997).
- Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13-18 (1997).
- Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31-36 (1998).
- Gelmini, S. et al. Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752-758 (1997).
- Emery, J. G. et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363-14367 (1998).
- Wallach, D. Placing death under control. *Nature* 388, 123-125 (1997).
- Collotta, F. et al. Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472-475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* **281**, 1305–1308 (1998).
23. Ashkenazi, A. & Chomow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195–200 (1997).
24. Marsters, S. *et al.* Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes¹. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters². Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains¹. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*^{3–8} is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP₂, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins⁹, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM⁶. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis⁵, the requirement for both subunits to be present for activity⁸, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer³. HisP has been purified and characterized in an active soluble form³ which can be reconstituted into a fully active membrane-bound complex⁸.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β -sheet ($\beta 3$ and $\beta 8$ – $\beta 12$) spans both arms of the L, with a domain of α - plus β -type structure ($\beta 1$, $\beta 2$, $\beta 4$ – $\beta 7$, $\alpha 1$ and $\alpha 2$) on one side (within arm I) and a domain of mostly α -helices ($\alpha 3$ – $\alpha 9$) on the

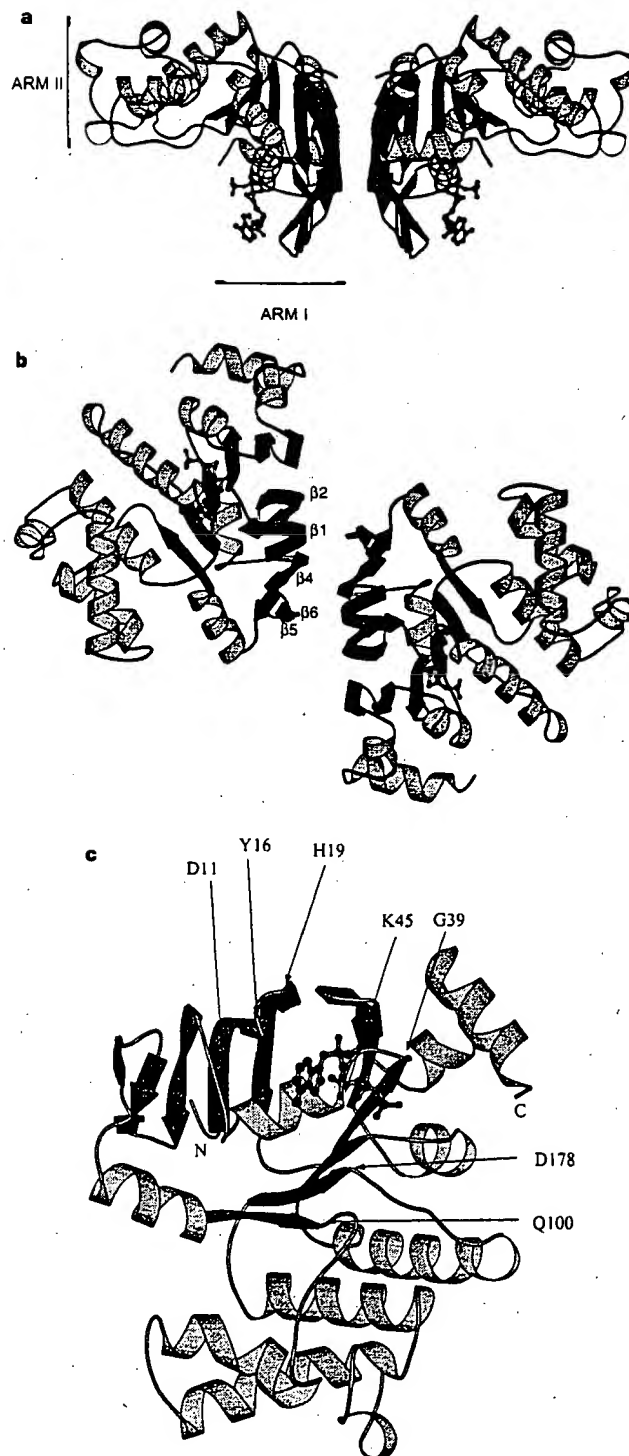


Figure 1 Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α -Helices are shown in orange and β -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The β -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a–c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT¹⁰. N, amino terminus; C, carboxyl terminus.

NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE^{1,2}, Martine OLIVI¹, Marie-Hélène CHAMPÈME², Dominique VIDAUD¹, Rosette LIDÉREAU² and Michel VIDAUD^{1*}

¹Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France

²Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.

© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmins) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuurin *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy

surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5' nuclease assay first described by Holland *et al.* (1991). The latter uses the 5' nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5' end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3' end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754. E-mail: mvidaud@teaser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (i.e., ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the R_n (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): (i) the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridization of both the primers and the probe is necessary to generate a signal; (ii) the C_t (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why C_t is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; (iii) use of C_t values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; (iv) The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; (v) the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; (vi) the assay is rapid, as results are available just one minute after thermal cycling is complete; (vii) the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erbB2*), as well as 2 genes (*alb* and *app*) located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

MATERIAL AND METHODS

Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguénin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

Real-time PCR

Theoretical basis. Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter C_t (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring C_t and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (i.e., lack

of extensive degradation) are both difficult to assess. We therefore also quantified a control gene (*alb*) mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (app, myc, ccnd1, erbB2)}}{\text{copy number of reference gene (alb)}}$$

Primers, probes, reference human genomic DNA and PCR consumables. Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAGency (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

Standard-curve construction. The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatak *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophoresed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/ μ l. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from 10^{-7} (10^5 copies of each gene) to 10^{-10} (10^2 copies). This series of diluted PCR products was aliquoted and stored at -80°C until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

PCR amplification. Amplification mixes (50 μ l) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes), $10\times$ TaqMan buffer (5 μ l), 200 μ M dATP, dCTP, dGTP, and 400 μ M dUTP, 5 mM MgCl_2 , 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at 50°C and 10 min at 95°C . Thermal cycling consisted of 40 cycles at 95°C for 15 s and 65°C for 1 min. Each assay included: a standard curve (from 10^5 to 10^2 copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

Equipment for real-time detection. The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates C_t and determines the starting copy number in the samples.

Determination of gene amplification. Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

RESULTS

To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the β -amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/ μ l. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as 10^2 copies or as many as 10^5 copies.

Copy-number ratio of the 2 reference genes (*app* and *alb*)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA

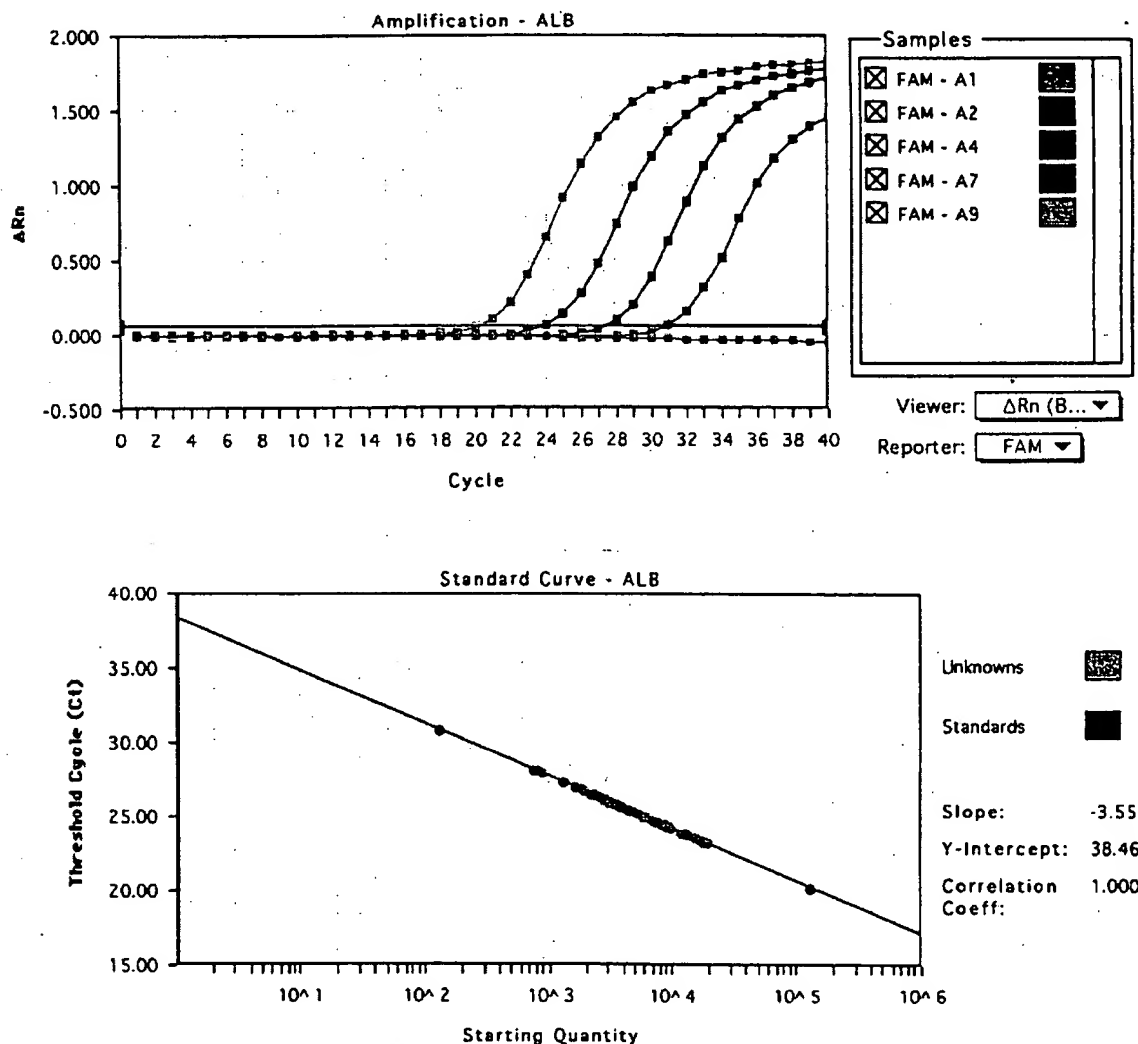


FIGURE 1 – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from 10^5 (A9), 10^4 (A7), 10^3 (A4) to 10^2 (A2) and a no-template control (A1). Cycle number is plotted vs. change in normalized reporter signal (ΔRn). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn). ΔRn represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles. ΔRn increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau. C_t (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number vs. C_t (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean 1.02 ± 0.21), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6, mean 1.06 ± 0.25), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

myc, *ccnd1* and *erbB2* gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean 0.84 ± 0.22) for *myc*, 0.7 to 1.6 (mean 1.06 ± 0.23) for *ccnd1* and 0.6 to 1.3 (mean 0.91 ± 0.19) for *erbB2*. Since N values for *myc*, *ccnd1* and *erbB2* in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

myc, *ccnd1* and *erbB2* gene dose in breast-tumor DNA

myc, *ccnd1* and *erbB2* gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erbB2* (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erbB2*, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erbB2* and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erbB2* in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erbB2* copy number ($N < 0.5$), suggesting that they bore deletions of the 17q21 region (the site of *erbB2*). No such decrease in copy number was observed with the other 2 proto-oncogenes.

Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erbB2* amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers ($N \geq 5$). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erbB2*) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive

reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

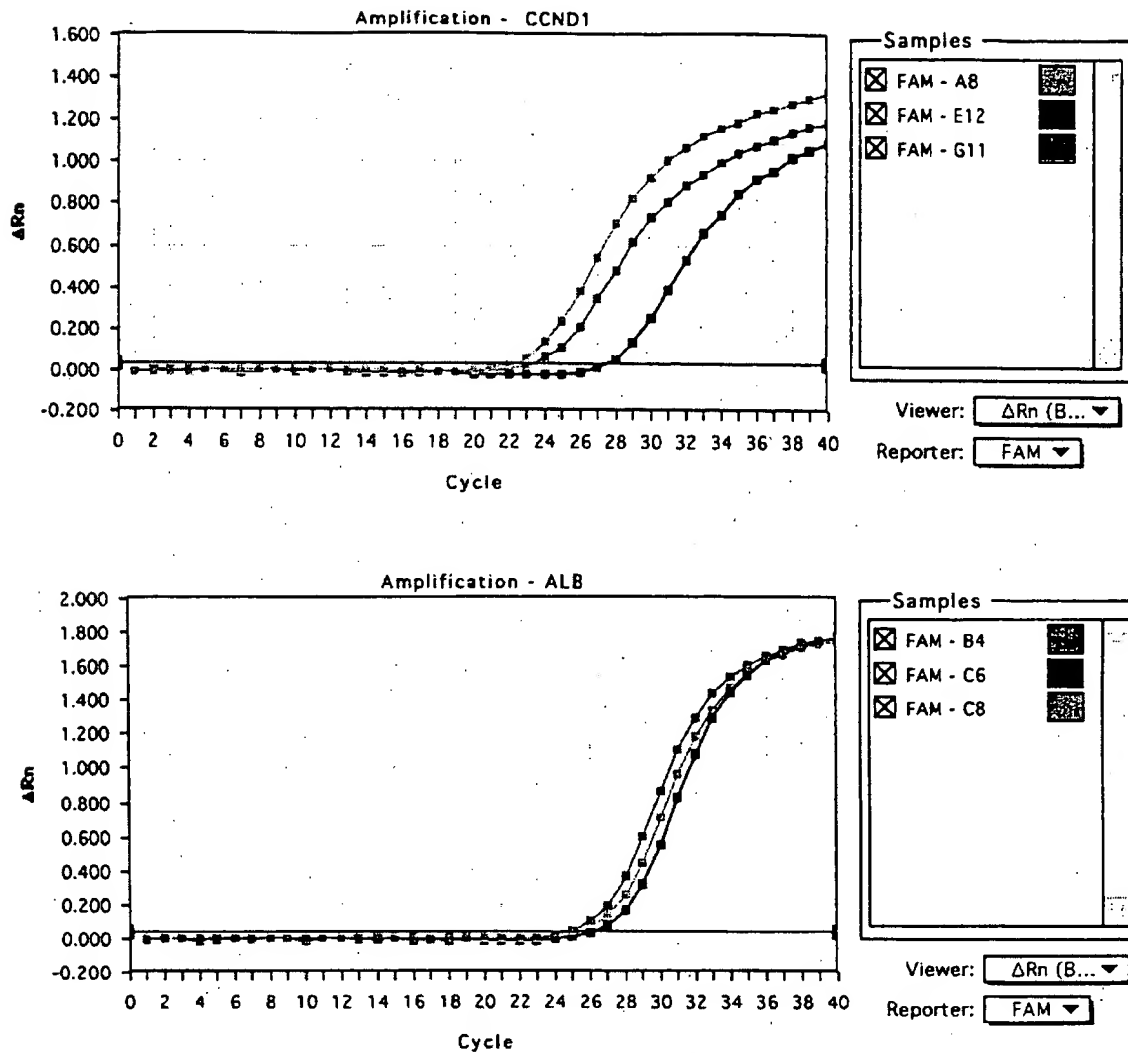
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on C_t values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables C_t to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the C_t value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the C_t ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

The only potential disadvantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that is cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (i) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (ii) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (iii) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erbB2* amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (iv) The maxima of *ccnd1* and *erbB2* over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I - DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erbB2* GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5-1.9	2-4.9	≥ 5
<i>myc</i>	0	97 (89.8%)	11 (10.2%)	0
<i>ccnd1</i>	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
<i>erbB2</i>	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)



Tumor	CCND1		ALB	
	C_t	Copy number	C_t	Copy number
■ T118	27.3	4605	26.5	4365
■ T133	23.2	61659	25.2	10092
■ T145	22.1	125892	25.6	7762

FIGURE 2 - *ccnd1* and *alb* gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the C_t of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erbB2* copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron

et al., 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erbB2* amplification in a small series of breast tumors ($n = 25$), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II - EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS¹

Tumor	<i>ccnd1</i>			<i>alb</i>			<i>Nccnd1/alb</i>
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525	4603	77	4223	4325	89	1.06
	4605			4365			
T133	4678	61100	1111	4387	10137	375	6.03
	59821			9787			
	61659			10092			
T145	61821	125392	3448	10533	7672	316	16.34
	128563			7321			
	125892			7762			
	121722			7933			

¹For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers (≥ 5 -fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erbB2*) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra

gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erbB2* (but not of the other 2 proto-oncogenes) in several tumors; *erbB2* is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

REFERENCES

- AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erbB2* gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, **64**, 291-297 (1995).
- BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTINGEN, H. and FOEKENS, J.A., *c-myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.*, **52**, 1107-1113 (1992).
- BIÈCHE, I. and LIDERAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, **14**, 227-251 (1995).
- BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., *c-myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, **51**, 687-691 (1992).
- CELI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (qd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, **21**, 304-310 (1994).
- COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, **57**, 4360-4367 (1997).
- DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erbB-2* and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, **40**, 271-281 (1996).
- GELMINI, S., ORIANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure *c-erbB-2* oncogene amplification. *Clin. Chem.*, **43**, 752-758 (1997).
- GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, **6**, 995-1001 (1996).
- HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, **6**, 986-994 (1996).
- HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, **88**, 7276-7280 (1991).
- KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, **91**, 2156-2160 (1994).
- LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, **21**, 3761-3766 (1993).
- LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, **93**, 125-128 (1990).
- MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, **330**, 1260-1266 (1994).
- PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of *HER-2/neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, **13**, 63-72 (1996).
- PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, **14**, 70-80 (1993).
- SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILEE, P., CORNELISSE, C.J., VAN DE VIJVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, **52**, 5229-5234 (1992).
- SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*, **235**, 177-182 (1987).
- SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDVOE, J., ULLRICH, A. and PRESS, M.F., Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*, **244**, 707-712 (1989).
- VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying *HER-2/neu* status in breast cancer. *Int. J. Cancer*, **65**, 129-133 (1996).

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.
App. No. : 09/903,925
Filed : July 11, 2001
For : SECRETED AND
TRANSMEMBRANE
POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan[®] PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth: 29 November, 1956
Address: 1456 Tarrytown Street, San Mateo, CA 94402
Phone: (650) 578-9199 (home); (650) 225-1853 (office)
Fax: (650) 225-6443 (office)
Email: aa@gene.com

Education:

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel
1986: Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986: Teaching assistant, undergraduate level course in Biochemistry
1985-1986: Teaching assistant, graduate level course on Signal Transduction
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,
with D. Capon
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,
Genentech, Inc.
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988: First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotropic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).
11. Peralta, E. Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine_{1B} receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT_{1B} receptors. *Nature* **360**, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).
26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3⁺ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* **5**, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT₂ receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* **92**, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal *E. coli* bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- κ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- α receptor 2. *J. Immunol.* **163**, 766-773 (1999).
52. J hnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563–571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* **29**, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* **7**, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* **166**, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* **7**, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* **2**, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* **276**, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* **8**, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, G., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* **170**, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahri, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzimology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoconjugates for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoconjugates: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoconjugate vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoconjugates: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of
Biochemistry, Michigan State University
East Lansing, Michigan

PUBLICATIONS:

1. **Polakis, P. G.** and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.
2. **Polakis, P.G.** and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.
3. **Polakis, P. G.** and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.
4. Uhing, R.J., **Polakis, P.G.** and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.
5. **Polakis, P.G.**, Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.
6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.
7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.
8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.
9. **Polakis, P.**, Weber, R.F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.
10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.
12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K⁺ Channels. **Cell.** 61, 769-776.
13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.
14. **Polakis, P.G.** Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.
15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812.
16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21^{ras} Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.
17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21^{rap1}. **Cell** 65, 1033-1042.
18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.
19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21^{ras}-dependent Inhibition of Muscarinic Atrial K⁺ Channel Currents. **Science** 255, 192-194.
20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P.** 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.
21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60^{v-src} Can Associate with and Phosphorylate the p21^{ras} GTPase Activating Protein. **Oncogene** 7, 389-394.
22. **Polakis P.** and McCormick, F. 1992 Interactions Between p21^{ras} Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34^{cdc2} Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21^{rap1A} and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell. Biol.** 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21^{ras} with GAP, exchange factors, and its biological effector target. **J. Biol Chem.** 268, 9157-9160.
31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., Polakis, P. and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. **Polakis, P.** 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α -catenin, β -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β -catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P.** 1997 Loss of β -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and **Polakis, P.** 1997 Induction of a β -catenin-LEF-1 complex by wnt-1 and transforming mutants of β -catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β -catenin and glycogen synthase kinase 3 β . **Current Biology** 8, 573-581.

56. **Polakis, P.** 1998 The oncogenic activation of β -catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and **Paul Polakis.** 1999 The F-box protein β -TrCP associates with phosphorylated β -catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β -catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P.**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwara, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β -catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.
- 61 **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32
- 62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.
65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.
66. **Polakis P**. 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.
67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.
69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.
70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.
71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman||, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37–45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, §AAROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ||UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, cyclin d1, ems1, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)¹ have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q–, 11p–, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary).

¹ The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

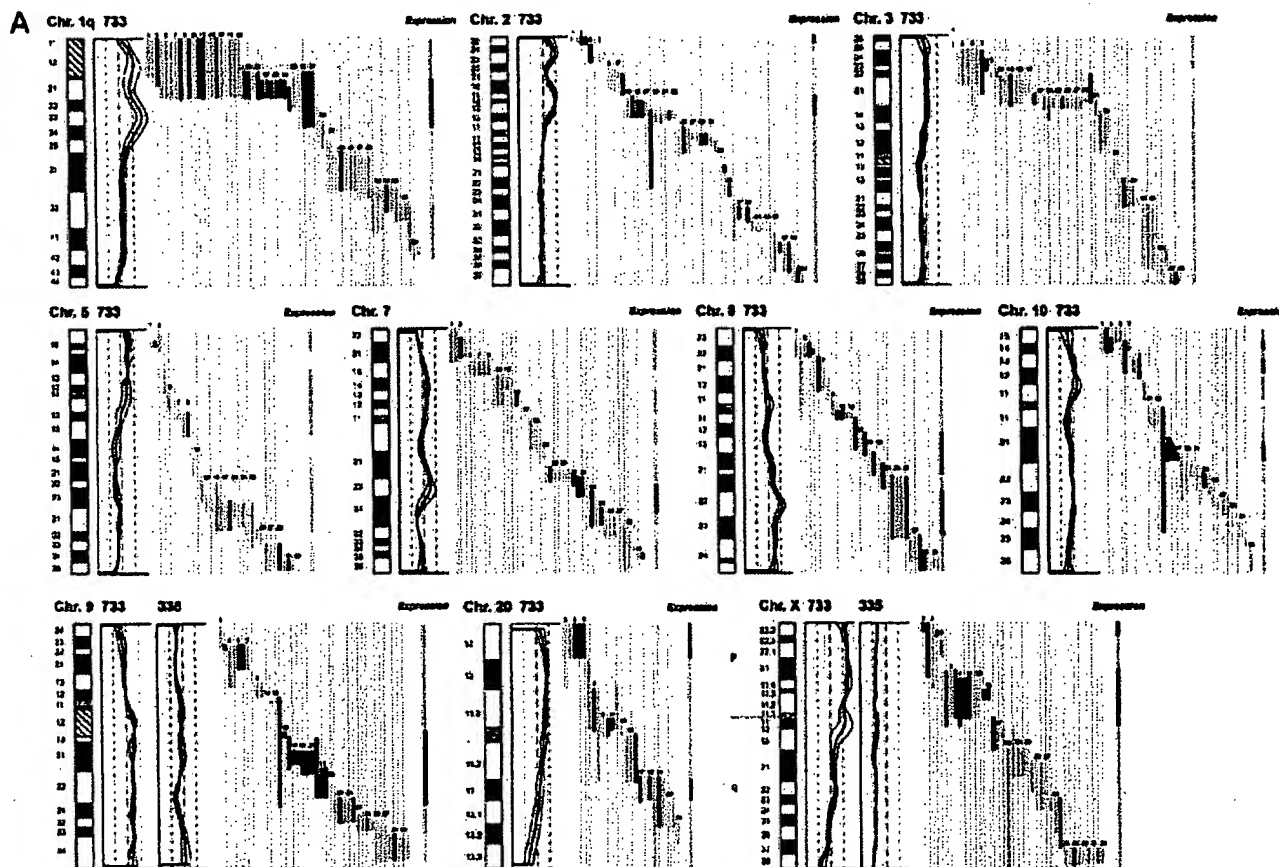


Fig. 1. DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. **A**, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. **B**, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a mean of four chromosomes and is surrounded by thin curves indicating one standard deviation. The central vertical line (broken) indicates a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at www.MDL.DK/sdata.html). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled Expression shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at -80°C . Total RNA was isolated using the RNeasy B RNA isolation method (WAK-Chemie Medical GmbH). poly(A)⁺ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript[®] choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscript[®] *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94°C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 \times SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95°C for 5 min, subsequently cooled to 40°C , and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40°C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 \times SSPE-T at 25°C followed by 4 washes in 0.5 \times SSPE-T at 50°C . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 $\mu\text{g}/\text{ml}$ (Molecular Probes) in 6 \times SSPE-T

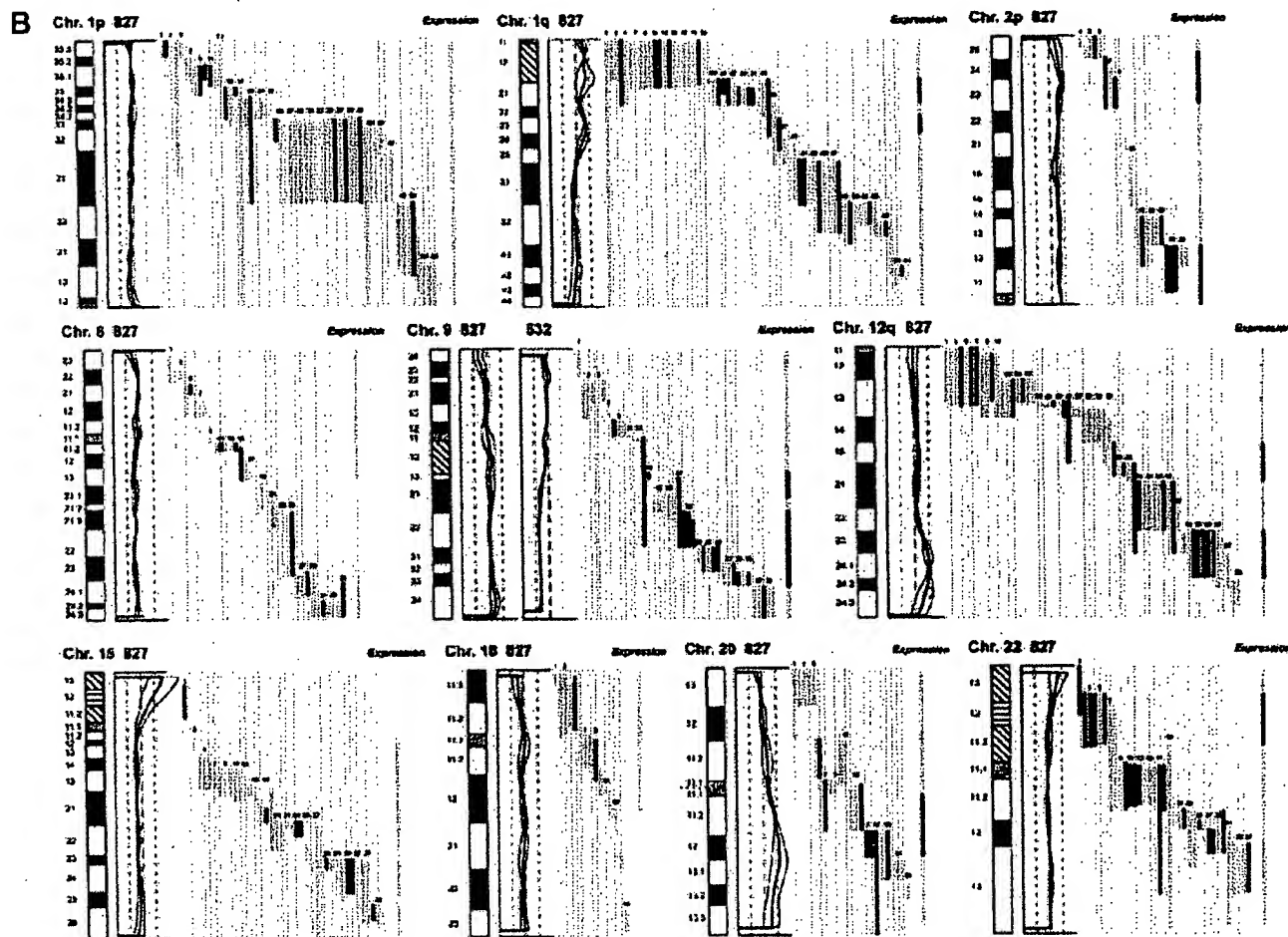


FIG. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20 μ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

Gene Copy Numbers, Transcripts, and Protein Levels

TABLE I
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335	Concordance	CGH alterations	Tumor 827 vs. 532	Concordance
	Expression change clusters			Expression change clusters	
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%
Expression change clusters	Tumor 733 vs. 335	Concordance	Expression change clusters	Tumor 827 vs. 532	Concordance
	CGH alterations			CGH alterations	
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

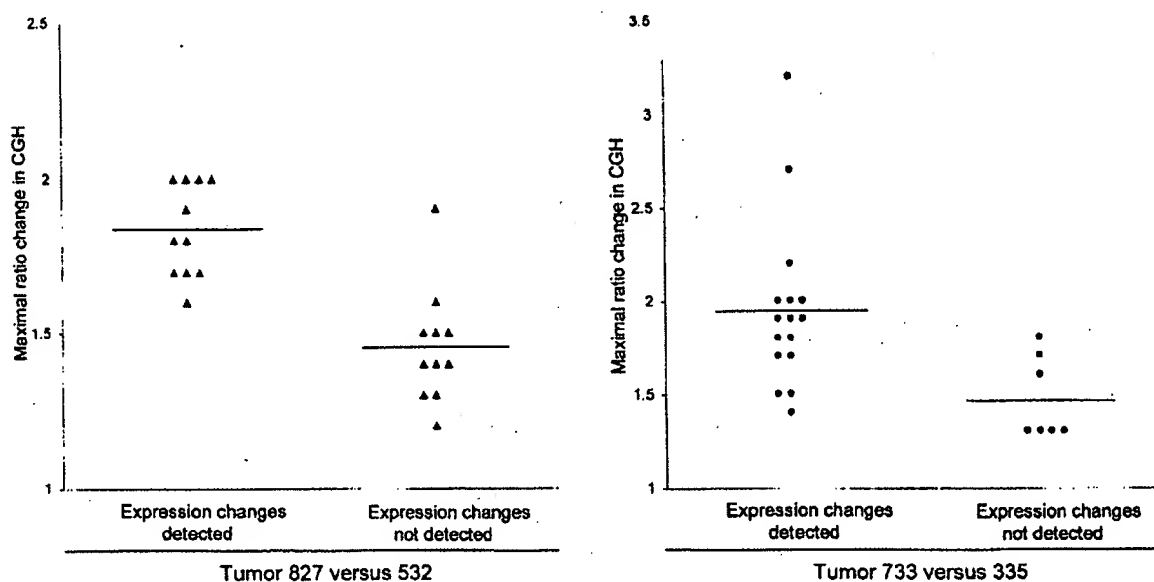


Fig. 2. Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring. The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (◆) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the right in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Figure 1 is a scatter plot showing the fold change in mRNA expression (Y-axis, logarithmic scale from 0.1 to 10) versus 2-D-gel position (X-axis). The X-axis is divided into three regions: Reduced protein, Unaltered protein, and Increased protein. A horizontal line at Y=1 indicates no change in mRNA expression. Data points are labeled A through L. Points A and B are in the Reduced protein region and show no change in mRNA expression. Points C and D are in the Reduced protein region and show a decrease in mRNA expression. Points E, F, and G are in the Unaltered protein region and show a decrease in mRNA expression. Points H, I, J, K, and L are in the Increased protein region and show an increase in mRNA expression.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

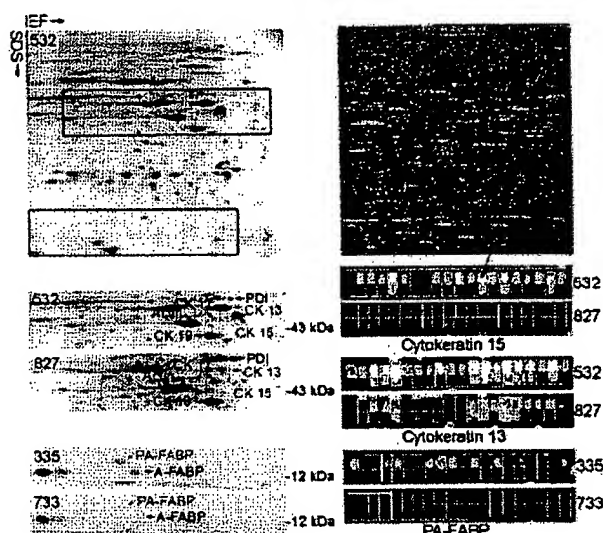


Fig. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs. The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration ^a	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres ^a	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up ^b	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

^a Abs, absent; Pres, present.

^b In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p–, 9q–, 1q+, Y– (2, 6), and in pT1 tumors, 2q–, 11p–, 11q–, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p– and 9q22–q33– and 9q– and Y–, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14–q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p–, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Nikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frønkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 396, 643–649.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46.
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157–161.
- Brungier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1998) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747–1753.
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459–1463.
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) *c-myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* 146, 1131–1139.
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860–2864.
- Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615–1621.
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167–174.
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 58, 3555–3560.
- Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345–351.
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87–93.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15, 1359–1367.
- Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167–177.
- Celis, J. E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782–4790.
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) In *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL.
- Ohlsson, R., Tycko, B., and Sapientza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435–438.
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118–2121.
- Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164–170.
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpu, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3889–3892.
- Cui, H., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469–4473.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251–254.
- Tsao, J., Yatabe, Y., Markl, I. D., Haiyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 26–32.
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632–10636.
- Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537.
- Ideker, T., Thorsson, V., Raniish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer^{1,2}

Elizabeth Hyman,³ Päivikki Kauraniemi,³ Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahoul, Olli-P. Kallioniemi, and Anne Kallioniemi⁴

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E.H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E.H., P.K., S.H., M.W., S.M., E.R., M.R., A.E., O.K., A.K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P.K., A.K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S.H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G.S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O.M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

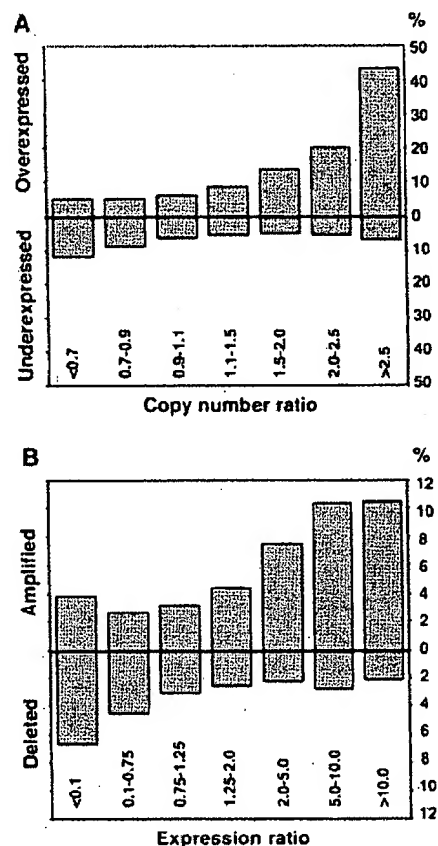


Fig. 1. Impact of gene copy number on global gene expression levels. A, percentage of over- and underexpressed genes (Y axis) according to copy number ratios (X axis). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). B, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7 .

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH⁵ (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

² Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

³ Contributed equally to this work.

⁴ To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkkelijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

⁵ The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

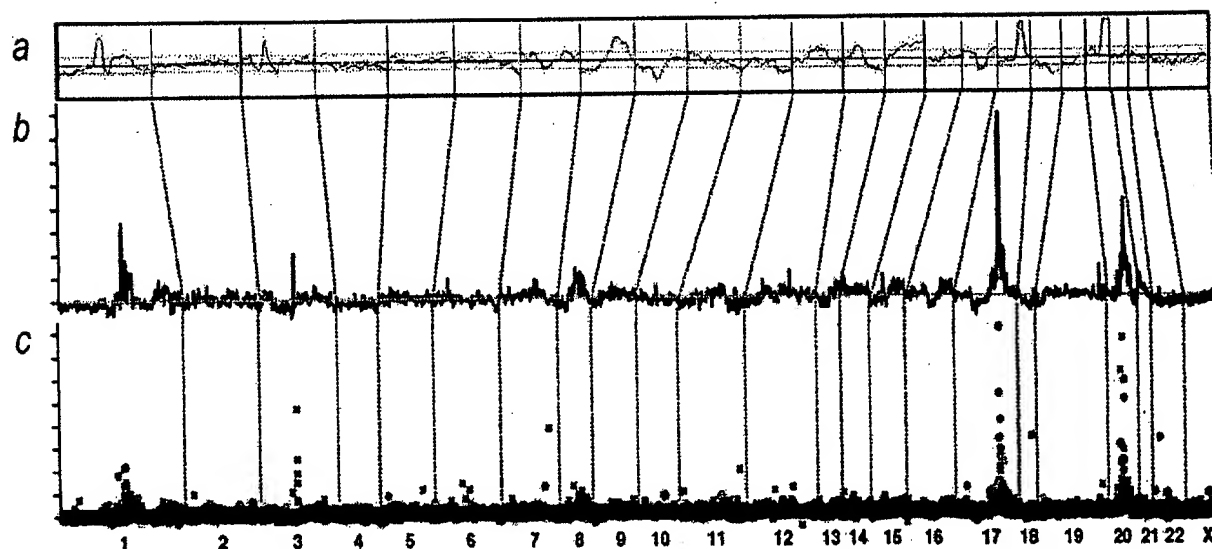


Fig. 2. Genomic-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with ± 1 SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B-C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright red dots indicate the upper 2%, and dark red dots, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosome numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μ g of reference RNA were labeled with Cy3-dUTP and 3.5 μ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1} , σ_{g1} and m_{g0} , σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α . A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.⁶ A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.⁷ The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

⁶ Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.

⁷ Internet address: www.genome.ucsc.edu.

Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

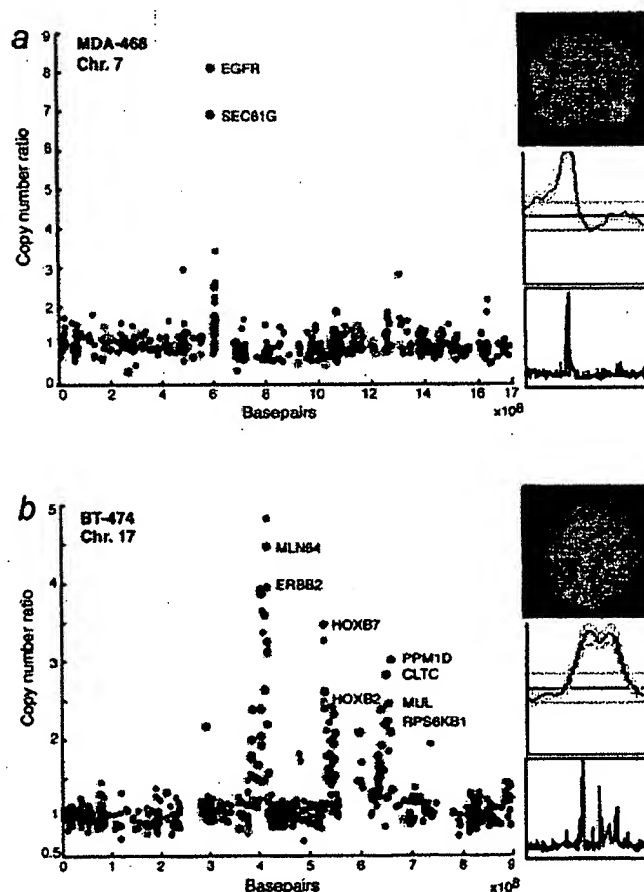


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

⁸ Internet address: <http://www.geneontology.org/>.



The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein *s6* kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531–537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503–511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536–540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747–752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822–826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869–10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125–138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32–40, 2001.
- Knuutila, S., Björkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457–460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639–645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevasian, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*, pp. 113–137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 41–46, 1999.
- Monni, O., Bärilund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711–5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364–374, 1997.
- Bärilund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Thorst, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein *S6* kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252–1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83–86, 2001.
- Kauraniemi, P., Bärilund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235–8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104–114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kakkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625–2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linchan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
25. Bärklund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

Protocol 12

RNA Common Reference Sets

Jonathan R. Pollack

Department of Pathology, Stanford University School of Medicine, Stanford, California 94305

To measure gene expression using microarrays, two different mRNA preparations are labeled, using different fluors, and cohybridized to a cDNA microarray. For each gene-specific element in the array, the ratio of intensity of the two fluors reflects the relative abundance of that particular mRNA in the two preparations.

The use of two-color labeling has the important property of robustly eliminating effects due to variation in the amount of DNA spotted on the arrays, variation in the stringency of the hybridization (salt concentration, temperature, and duration of hybridization), and variation in the local concentration of labeled sample during hybridization. Any changes in these parameters should affect the numerator and denominator of fluorescence ratios equally and thus would not influence measured fluorescence ratios.

In studies that include many different RNA preparations, it becomes impractical to hybridize all pair-wise combinations of samples. The solution is to hybridize each of the experimental samples against the same common reference RNA preparation, and to calculate the level of each mRNA relative to the reference. The relative abundances of mRNAs among the different experimental samples are then calculated indirectly, from the ratio of ratios.

For many experiments, a "natural" RNA reference sample is available, for example, when quantifying changes in mRNA abundance over time (reference = time zero sample) or after treatment with a drug (reference = control unexposed to drug). However, in certain instances, such as in the measurement of mRNA abundances in different tumor samples, no natural reference exists. In such cases, an mRNA "common reference set" can be used (Eisen and Brown 1999). If each test (i.e., tumor) sample is compared to the same common reference set, then the relative intensity of gene expression among the samples can be calculated. In this type of experiment, the common reference set becomes the denominator for all measured fluorescence ratios.

An ideal common reference set should hybridize with sufficient signal intensity to all or nearly all of the DNA elements in the array, should be easy and cost-effective to prepare, and should be easily reproduced to allow comparison of samples over time and between locations. Note that the common reference set need not have any particular biological purpose or property and need not include RNA from every cell type under investigation. However, a good reference set should generate a readable signal from most elements on the array. In addition, for the reference set to be useful to a broad group of researchers, it should be constructed from validated sources that are readily available. In practical terms, a common reference comprising pooled RNA from stable lines of cultured cells meets many of these criteria.

This protocol outlines the preparation of an RNA common reference set used extensively by the cDNA microarray community at Stanford University (Perou et al. 1999, 2000; Ross et al. 2000). This reference set is produced from 11 different established cell lines (see Table 3-6), many of which are available from the American Type Culture Collection. These cell lines

are easy to grow and, in combination, hybridize to a large majority of arrayed cDNA elements. Although the optimal number of cell lines used in a reference set has not been established experimentally, it is perhaps important not to use too many different cell types. The pooled RNA may dilute to the point of invisibility low-abundance transcripts that are expressed in only one or a few of the cell lines in the set.

Most laboratories with the infrastructure required for DNA microarrays should be able to reproduce the Stanford common reference set. In general, the work load involved in producing a batch of the common reference set typically could require a period of 2–3 months of full-time work to grow up the 11 cell lines and purify sufficient RNA for several thousand array hybridizations. However, for some investigators, a commercial reference set may be the best option. Stratagene Inc. sells a common reference set of RNA extracted from ten human cell lines (Novoradovskaya et al. 2000). However, unlike the Stanford set, few details are available about the cell lines and their provenance.

Both the Stanford and the Stratagene common reference sets allow rational comparison of microarray results within and between laboratories. Other possible reference sets, less commonly used, that may be able to fulfill the same purpose include labeled genomic DNA or RNA generated by transcription of a pool of cDNAs *in vitro*.

WHY USE A COMMON REFERENCE SET?

Consider an experiment whose goal is to identify genes that are expressed at different levels in tumors and normal cells. Two experimental designs are possible:

- Tumor versus tumor versus normal

Samples: T1, T2, N1, N2

Options for comparison:

T1 vs. N1, T2 vs. N2, T1 vs. N2, T2 vs. N1, T1 vs. T2, N1 vs. N2

This matrix provides a comparison of all possible combinations and has the advantage that questions are addressed by direct comparison: How does one tumor compare with another? How does tumorous tissue compare with normal? How does normal compare with normal? However, this approach is only practical when the number of samples is very small, as the number of combinations increases exponentially with the number of samples.

- Tumor versus reference, normal versus reference

Samples: T1, T2, N1, N2, R

Options for comparison:

T1 vs. R, T2 vs. R, N1 vs. R, N2 vs. R

Here, all samples are tested against a common reference RNA and comparison is made indirectly with respect to the reference. The number of combinations is manageable as they increase linearly with the number of samples. In addition, comparisons can be made between experiments and investigators who use a common reference. A potential disadvantage is that inferences are made indirectly between samples, via the reference, potentially reducing the resolving power of the experiment. The ideal common reference therefore should be widely accessible, available in unlimited amounts, and should provide a signal over a wide range of genes. These issues are considered in more detail in Section 7.

Figure 3-12 illustrates the differences in these two experimental designs.

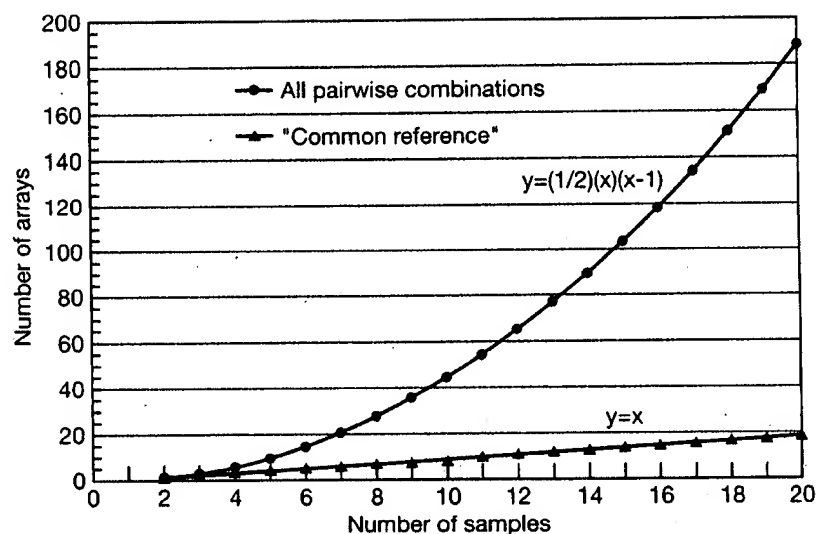


FIGURE 3-12. Differences in performing all pairwise combinations vs. using a common reference.

MATERIALS

CAUTION: Please see Appendix 3 for appropriate handling of materials marked with <!.>.

Buffers and Solutions

Please see Appendix 2 for components of stock solutions, buffers, and reagents. Dilute stock solutions to the appropriate concentrations.

ATRA (all-*trans* retinoic acid; R 2625, Sigma)
Prepare a 10 mM stock in DMSO.

Cell culture media

RPMI 1640 (GIBCO or equivalent)
10% fetal bovine serum (Hyclone or equivalent)
penicillin and streptomycin (1x concentration, GIBCO)
Phosphate-buffered saline (PBS), lacking Ca^{2+} and Mg^{2+}
1x Trypsin-EDTA

Gels

Agarose gel (1%) containing 2.2 M formaldehyde <!.>
For details, please see Sambrook and Russell (2001).

Special Equipment

FastTrack 2.0 mRNA isolation kit (Invitrogen)
Humidified incubator with an atmosphere of 5% CO_2
Laminar flow hood for passaging and harvesting of cells
Roller bottles and appropriate incubator (alternative to tissue culture dishes)
Tissue culture dishes (15 cm)
UV spectrophotometer

TABLE 3-6. Cell Lines Used in Stanford "Common RNA Reference"

Name	Description	Growth properties	ATCC catalog No. or Reference
MCF7	breast adenocarcinoma-derived cell line	adherent	ATCC HTB-22
Hs578T	breast adenocarcinoma-derived cell line (stromal-like)	adherent	ATCC HTB-126
NTERA2	teratoma-derived cell line	adherent	ATCC CRL-1973
Colo205	colon tumor-derived cell line	mixed	ATCC CCL-222
OVCAR-3	ovarian tumor-derived cell line	adherent	ATCC HTB-161
UACC-62	melanoma-derived cell line	adherent	Stinson et al. (1992)
MOLT-4	T cell leukemia-derived cell line	suspension	ATCC CRL-1582
RPMI 8226	multiple myeloma-derived cell line	suspension	ATCC CCL-155
NB4+ATRA	acute promyelocytic leukemia-derived cell line	suspension	Lanotte et al. (1991)
SW872	liposarcoma-derived cell line	adherent	ATCC HTB-92
HepG2	liver tumor-derived cell line	adherent	ATCC HB-8065

Cells and Tissues

Cell lines

Please see Table 3-6 for a listing and description of cell lines used in the Stanford common reference set of RNAs. Of the 11 cell lines, 9 are currently available from the American Type Culture Collection (www.atcc.org). However, because NB4 and UACC-62 are not at present distributed by the ATCC, it is impossible to replicate the Stanford Common Reference Set from publicly available sources. We therefore recommend that HL-60 cells (ATCC CCL-240) be substituted for NB4 cells and that the SK-MEL-2 melanoma cell line (ATCC HTB-68) be substituted for UACC-62 cells. Because the aim should be to preserve the "commonality" of the reference set as far as possible, other substitutions by individual laboratories are discouraged.

METHOD

Growing Cell Lines

The 11 cell lines listed in Table 3-6 are easy to maintain and grow in either monolayer or suspension culture. This protocol is designed for cells growing in culture dishes. If suspension cultures are grown in flasks or roller bottles, adjust the protocol accordingly.

1. Maintain the 11 cell lines in RPMI 1640 supplemented with 10% fetal calf serum in 15-cm tissue culture dishes at 37°C in an atmosphere of 5% CO₂ in air.

Cells should be passaged when approaching confluence at ~1:5 dilution using standard trypsinization procedures for adherent lines.

Record passage numbers of cell lines in a log book or database. Store master cultures of early passages in liquid nitrogen in small aliquots in medium containing 10% DMSO.

Discard cell lines after ten passages in cell culture and replace them from frozen master cultures.

Wherever possible, use the same batch of fetal calf serum to grow the 11 cell lines. Serum is probably a significant source of batch-to-batch variation in reference RNAs. Record the batch numbers of serum and medium used for each passage of each cell line.

2. Expand the cell numbers so that 10–20 15-cm tissue culture dishes of each cell line can be grown.

3. Change the media 48 hours before harvesting the cells for isolation of mRNA.

Media changes produce alterations in the pattern and intensity of gene expression. To minimize variation between preparations of common reference RNAs, establish standardized protocols for growing and refeeding cells and follow them faithfully.

4. Add *all-trans* retinoic acid (ATRA) to a final concentration of 10 μ M to the culture medium of NB4 cells (or substituted HL-60 cells) ~48 hours before harvesting the cells.
NB4 and HL-60 cells are derived from cases of human promyelocytic leukemia. Addition of ATRA induces, over the course of 48 hours, the expression of genes specific to granulocytic differentiation with concomitant maturation of the cells.
5. Harvest the cells for mRNA isolation at ~80% confluence.
Patterns of gene expression change as cell cultures approach confluence. To minimize variation among preparations of common reference RNAs, standardize protocols for growing and harvesting cells and follow them faithfully.

Preparation of mRNA

6. Prepare mRNA from the cell lines using, for example, the FastTrack 2.0 (Invitrogen) mRNA isolation kit.
 - a. Remove the media from the culture by aspiration.
 - b. Lyse the cells on the surface of the dishes using the lysis buffer supplied by the manufacturer.
 - c. Isolate the mRNA by batch chromatography on oligo(dT)-cellulose.
A number of methods can be used to prepare either total RNA or mRNA from cell lines. The Stanford group routinely uses the FastTrack 2.0 (Invitrogen) mRNA isolation kit. The FastTrack kit yields mRNA of high quality for labeling and hybridization to DNA microarrays, although less expensive alternatives are likely to be equally effective. An outline of the procedure is given here; for details please see Protocol 6.
7. Determine the yield and purity of mRNA by UV spectrophotometry or staining with RiboGreen (please see the information panel on QUANTIFICATION OF RNA).
8. Assess the quality and integrity of the mRNA by denaturing gel electrophoresis through formaldehyde-agarose gels (please see information panel on MAMMALIAN, PLANT, AND BACTERIAL RNAs).
9. If the quantity and quality of the isolated mRNA are satisfactory, combine equal amounts (weight) of mRNA from each cell line and mix well. Distribute the pooled RNA mixture into small aliquots (10–20 μ g), and store them at -80°C or as described in the information panel on STORAGE AND RECOVERY OF RNA.

CALIBRATING DIFFERENT BATCHES OF COMMON REFERENCE RNA

Although every effort is taken to grow the cell lines and isolate the mRNA in a uniform and standardized manner, even slight differences in the treatment of cells, or in the quantification of mRNA, may change the abundances of mRNA species in the reference set. To correlate results obtained from experiments using different batches of the reference sets, it may be necessary to quantify and correct for batch-to-batch variation. This correction may be achieved by labeling different batches of reference sets with Cy3 and Cy5 and hybridizing them in pair-wise crosses to DNA microarrays. After this experiment has been repeated several times, accurate batch-specific "calibration" coefficients can be calculated for each gene on the array. These coefficients may be used to correct for systematic differences caused by batch-to-batch variation in the common reference sets.



TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.¹ Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.²

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.³

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.⁴ HER-2/neu status may be particularly important to establish in women with small (≤ 1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.⁵ Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

88271×2 Molecular cytogenetics, DNA probe, each

88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells

88291 Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest®. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion™ HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The Pathvysion™ kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-s/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

**PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY**

*For more information, please contact
your local representative.*

Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. *Curr Opin Oncol* 2000, 12:82-88 © 2000 Lippincott Williams & Wilkins, Inc.

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1515 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2580; fax: 713-792-4094; e-mail: ssen@mdanderson.org

Current Opinion in Oncology 2000 12:82-88

Abbreviations

CGH . comparative genomic hybridization
CHE Chinese hamster embryo cells
FISH fluorescence *in situ* hybridization
HPRC hereditary papillary renal carcinoma
ISH *in situ* hybridization

ISSN 1040-8746 © 2000 Lippincott Williams & Wilkins, Inc.

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4•]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage

compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during

tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of 10^{-2} per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in*

vitro model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

Aneuploidy versus somatic gene mutation in cancer

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ($\sim 10^{-7}$ per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary

centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the *Xenopus* homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein Xif5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in *p53*-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (v-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the v-securin may contribute to generation of malignant tumors due to

Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis

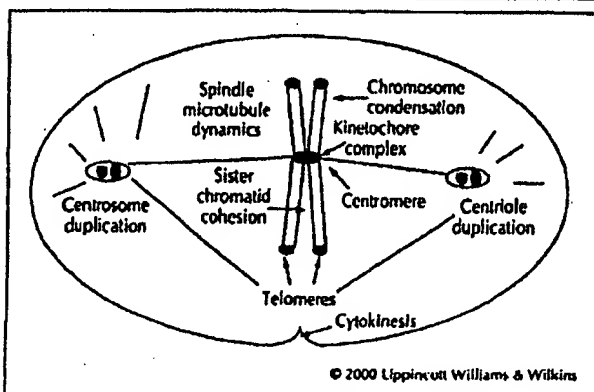


Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.

chromosome gain or loss produced by errors in chromatid separation [43*].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45*] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possibility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- Of special interest
- Of outstanding interest

- 1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley Liss Inc., 1995.
- 2 Nowell PC: The clonal evolution of tumor cell populations. *Science* 1976, 194:23-28.
- 3 Tomlinson IP, Novelli MR, Bodmer WF: The mutation rate and cancer. *Proc Natl Acad Sci USA* 1996, 93:14800-14803.
- 4 Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396:643-649.
- An excellent review on the significance and possible mechanisms of genetic instability in cancer.
- 5 Heppner GH, Miller FR: The cellular basis of tumor progression. *Int Rev Cytol* 1998, 177:1-66.
- 6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. *Diag Cytopath* 1998, 18:18-23.
- 7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. *Oncology* 1996, 10:887-890.
- 8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. *Br J Biomed Sci* 1997, 54:140-148.
- 9 Fletcher JA: Renal and bladder cancers. In: *Human Cytogenetic Cancer Markers*. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169-202.
- 10 Zhuang Z, Park WS, Peck S, Schmidt L, Vortmeyer AO, Pak E, et al: Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nat Genet* 1998, 20:66-69.
- 11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999, 86:947-950.
- 12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. *Anal Quant Cytol Histo* 1999, 21:85-88.
- 13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JL, Lopez A, et al: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. *Anal Cell Path* 1998, 18:223-231.
- 14 DeAngelis PM, Clausen OP, Schjolberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

88 Cancer biology

- associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999, 80:528-535.
- 15 Malkhosyan S, Yasuda J, Scolio JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (aneuploidy) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. *Proc Natl Acad Sci (USA)* 1998, 95:10170-10175.
 - 16 Ai H, Barrera JE, Pan Z, Meyers AD, Varela-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. *Mol Res* 1999, 439:223-232.
 - 17 Barrera JE, Ai H, Pan Z, Meyers AD, Varela-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. *Arch Otolaryngol Head Neck Surg* 1998, 124:847-851.
 - 18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc Natl Acad Sci USA* 1998, 95:7081-7084.
 - 19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. *Cytometry* 1999, 34:287-293.
 - 20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett esophagus. *Nat Genet* 1999, 22:106-109.
 - 21 Sturges CD, Caraway NP, Johnston DA, Sherman SJ, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. *Cancer* 1999, 87:155-160.
 - 22 Hemmer S, Waserius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. *Br J Cancer* 1998, 78:1012-1017.
 - 23 Mendelin J, Grayson M, Wallis T, Vischer DW: Analysis of chromosome aneuploidy in breast carcinoma: progression by using fluorescence in situ hybridization. *Lab Inv* 1999, 79:387-393.
 - 24 Bulten J, Poddighe PJ, Robben JC, Gammink JH, deWilde PC, Hanselaar GAGM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. *Am J Pathol* 1998, 152:495-503.
 - 25 Krauter J, Ganser A, Bergmann L, Raghavacher A, Hoelzer D, Libbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. *Ann Hematol* 1999, 78:285-289.
 - 26 Van Den Neste E, Louvieux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. *Br J Hematol* 1999, 105:268-270.
 - 27 Namba M, Mihara K, Fushimi K: immortalization of human cells and its mechanisms. *Crit Rev Oncog* 1996, 7:19-31.
 - 28 Li R, Yerganian G, Duesberg P, Kraemer A, Weller A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc Natl Acad Sci USA* 1997, 94:14806-14811.
 - 29 Liu P, Zheng H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* 1998, 150:1155-1168.
 - 30 Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survival. *Nature* 1998, 396:580-584.
 - 31 Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. *Nature* 1997, 386:823-827.
Demonstrates chromosomal instability in aneuploid colorectal tumor cells.
 - 32 Duesberg P, Rausch C, Raznick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, 95:13892-13897.
Correlates aneuploidy and transformation in *in vitro* grown CHE cells.
 - 33 Boveri T: Zur Frage der Entstehung maligner Tumoren. Jena, Verlag von Gustav Fischer, 1914.
 - 34 Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? *Nat Biotech* 1998, 16:137-138.
Discusses the significance of aneuploidy and gene mutations in cancer.
 - 35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. *Nature* 1998, 392:223-224.
 - 36 Pihan GA, Dossey SJ: The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999, 9:289-302.
Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegregation in cancer.
 - 37 Pihan GA, Purohit A, Wallace J, Knecht H, Ubeda B, Queensberry P, Dossey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58:3974-3985.
 - 38 Lingle WJ, Lutz WH, Ingle JN, Mahle NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci USA* 1998, 95:2950-2955.
 - 39 Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20:189-193.
Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.
 - 40 Bischoff JR, Anderson L, Shu Y, Morse K, Ng I, Chan CS, et al.: A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17:3052-3065.
Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.
 - 41 Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C: The xenopus laevis aurora related protein kinase pEg2 associates with and phosphorylates the kinesin related protein X1Eq5. *J Biol Chem* 1999, 274:15005-15013.
 - 42 Zimmerman W, Sparks C, Dossey S: Amorphous no longer: the centrosome comes into focus. *Curr Opin Cell Biol* 1998, 11:122-128.
 - 43 Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999, 285:418-421.
Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.
 - 44 Hardwick KG: The spindle checkpoint. *Trends Genet* 1998, 14:1-4.
 - 45 Cahill DP, Lengauer C, Yu J, Riggins GJ, Wilson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. *Nature* 1998, 392:300-303.
Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.
 - 46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 1998, 33:81-91.
 - 47 Xu X, Weaver Z, Unke SP, U C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. *Mol Cell* 1999, 3:389-395.

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA^{*†}, TODD A. SWANSON^{*}, JAMES W. WELSH^{*}, MARGARET A. ROY[‡], DAVID A. LAWRENCE^{*}, JAMES LEE[‡], JENNIFER BRUSH[‡], LISA A. TANEYHILL[§], BETHANNE DEUEL[‡], MICHAEL LEW[¶], COLIN WATANABE^{||}, ROBERT L. COHEN^{*}, MONA F. MELHEM^{**}, GENE G. FINLEY^{**}, PHIL QUIRKE^{††}, AUDREY D. GODDARD[‡], KENNETH J. HILLAN[¶], AUSTIN L. GURNEY[‡], DAVID BOTSTEIN^{‡,‡‡}, AND ARNOLD J. LEVINE[§]

Departments of ^{*}Molecular Oncology, [‡]Molecular Biology, [§]Scientific Computing, and [¶]Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; ^{**}University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ^{††}University of Leeds, Leeds, LS29JT United Kingdom; ^{‡‡}Department of Genetics, Stanford University, Palo Alto, CA 94305; and [§]Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 β (GSK-3 β) resulting in an increase in β -catenin levels. Stabilized β -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β -catenin levels (9). APC is phosphorylated by GSK-3 β , binds to β -catenin, and facilitates its degradation. Mutations in either APC or β -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0
PNAS is available online at www.pnas.org.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

[†]To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μ g of poly(A)⁺ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μ g of poly(A)⁺ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

In Situ Hybridization. ³³P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{-\Delta\Delta C_t}$ where ΔC_t represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The δ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of \approx 40,000 (*M*, 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of \approx 27,000 (*M*, 27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

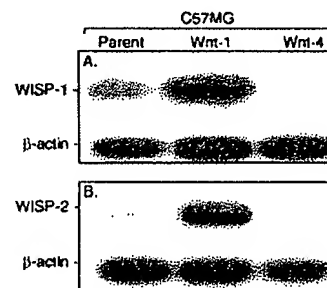


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)⁺ RNA (2 μ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human β -actin probe.

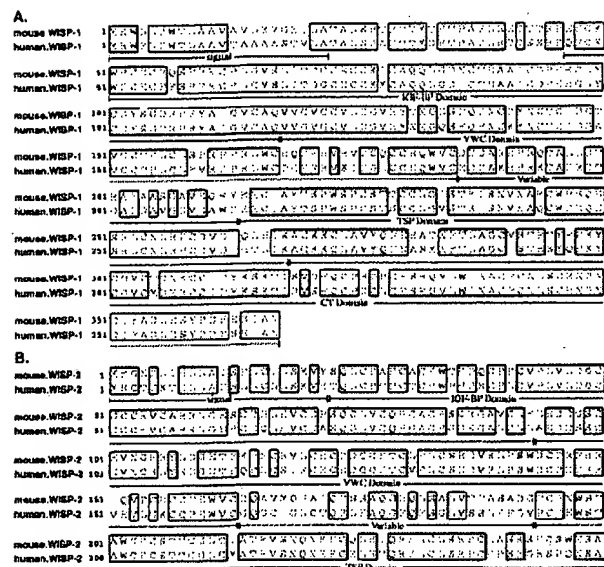


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

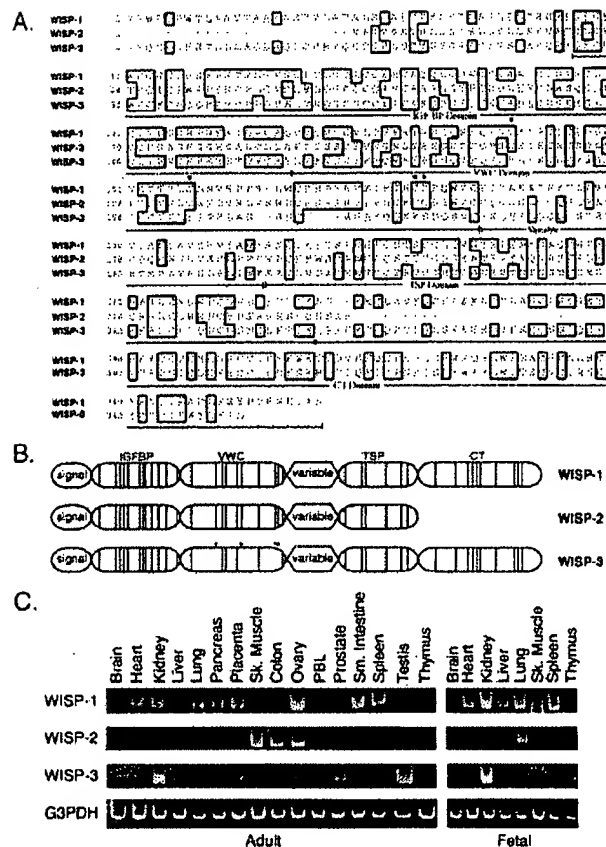


FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSXCSXCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

In Situ Localization of *WISP-1* and *WISP-2*. Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

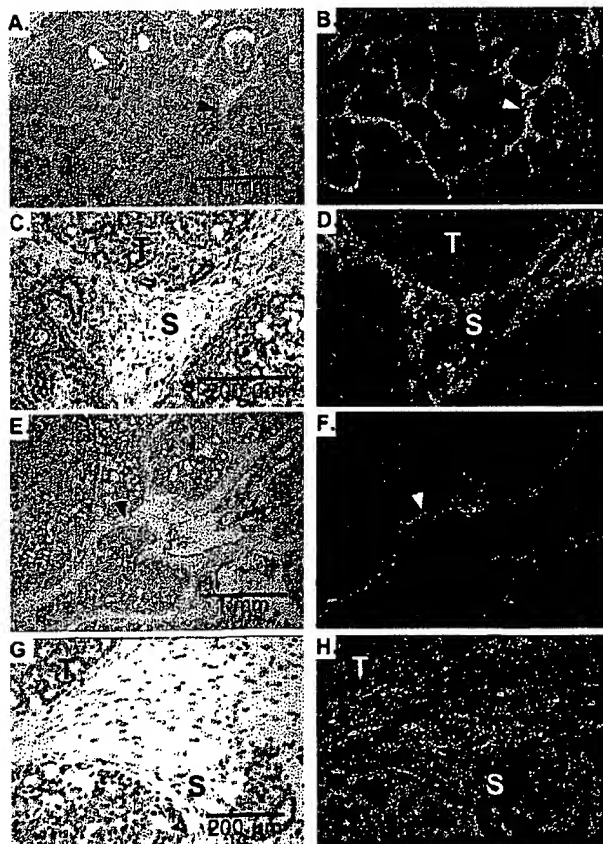


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

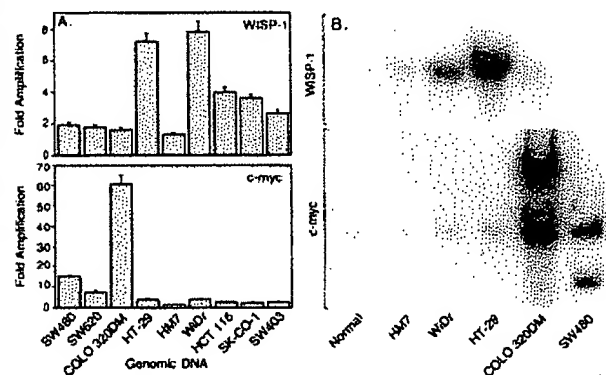


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10 μ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

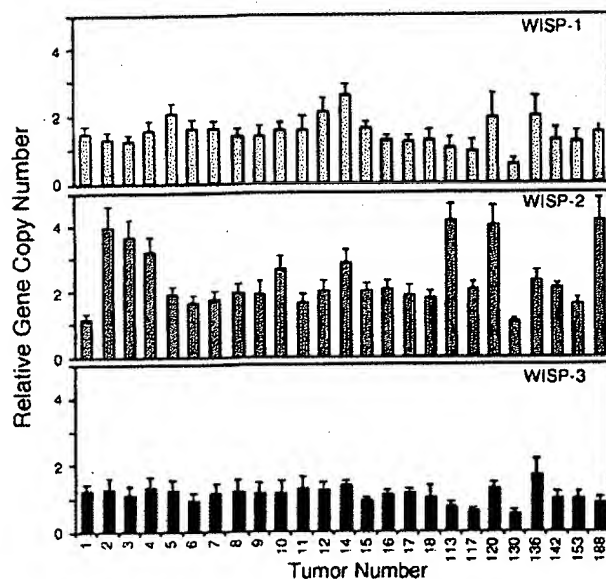


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

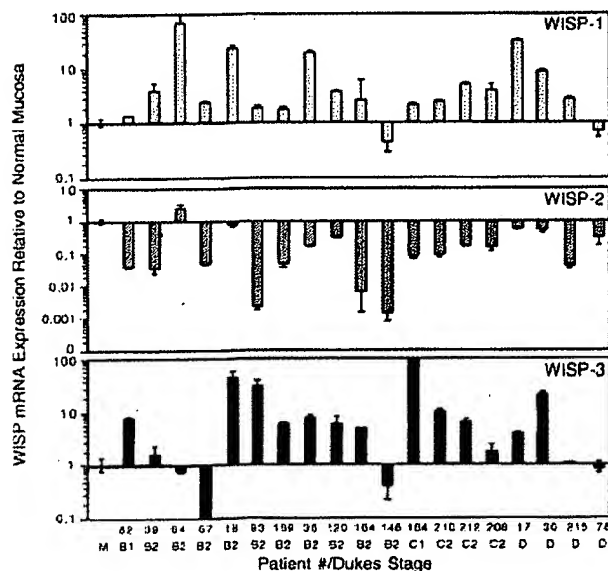


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means \pm SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- β , platelet-derived growth factor, and nerve growth factor, which contain a cysteine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- β 1, which is the stimulus for stromal proliferation (34). TGF- β 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of WISP-1 and WISP-2 in the stromal cells of breast tumors supports this paracrine model.

An analysis of WISP-1 gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human WISP-2 was localized to chromosome 20q12-20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36-38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for WISP-2 may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of WISP-2, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which WISP-2 RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of WISP-2 in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the WISP genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the WISPs. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of WISPs as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the WISPs in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* **11**, 3286-3305.
- Dale, T. C. (1998) *Biochem. J.* **329**, 209-223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* **31**, 99-109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* **39**, 233-240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* **55**, 619-625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* **10**, 182-187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* **86**, 391-399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* **18**, 1248-1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3046-3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* **281**, 1509-1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6025-6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* **46**, 1001-1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* **14**, 6278-6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* **8**, 1349-1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* **187**, 289-296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* **18**, 6131-6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* **8**, 171-179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* **16**, 1326-1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6355-6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* **9**, 2729-2732.
- Bork, P. (1993) *FEBS Lett.* **327**, 125-130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12981-12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiat, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* **12**, 10-21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 19514-19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* **265**, 2852-2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* **113**, 195-205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* **7**, 2529-2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* **57**, 109-111.
- Meesse, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* **1**, 88-94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* **4**, 435-449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **276**, 1268-1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269-291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* **273**, 3090-3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* **29**, 153-161.
- Wernert, N. (1997) *Virchows Arch.* **430**, 433-443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., et al. (1994) *Cancer Res.* **54**, 4257-4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* **6**, 187-194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., et al. (1998) *EMBO J.* **17**, 3052-3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **275**, 1787-1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* **1**, 169-176.

Correlation between Protein and mRNA Abundance in Yeast

STEVEN P. GYGI, YVAN ROCHON, B. ROBERT FRANZA, AND RUEDI AEBERSOLD*

Department of Molecular Biotechnology, University of Washington, Seattle, Washington 98195-7730

Received 5 October 1998/Returned for modification 11 November 1998/Accepted 2 December 1998

We have determined the relationship between mRNA and protein expression levels for selected genes expressed in the yeast *Saccharomyces cerevisiae* growing at mid-log phase. The proteins contained in total yeast cell lysate were separated by high-resolution two-dimensional (2D) gel electrophoresis. Over 150 protein spots were excised and identified by capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS). Protein spots were quantified by metabolic labeling and scintillation counting. Corresponding mRNA levels were calculated from serial analysis of gene expression (SAGE) frequency tables (V. E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler, *Cell* 88:243-251, 1997). We found that the correlation between mRNA and protein levels was insufficient to predict protein expression levels from quantitative mRNA data. Indeed, for some genes, while the mRNA levels were of the same value the protein levels varied by more than 20-fold. Conversely, invariant steady-state levels of certain proteins were observed with respective mRNA transcript levels that varied by as much as 30-fold. Another interesting observation is that codon bias is not a predictor of either protein or mRNA levels. Our results clearly delineate the technical boundaries of current approaches for quantitative analysis of protein expression and reveal that simple deduction from mRNA transcript analysis is insufficient.

The description of the state of a biological system by the quantitative measurement of the system constituents is an essential but largely unexplored area of biology. With recent technical advances including the development of differential display-PCR (21), of cDNA microarray and DNA chip technology (20, 27), and of serial analysis of gene expression (SAGE) (34, 35), it is now feasible to establish global and quantitative mRNA expression profiles of cells and tissues in species for which the sequence of all the genes is known. However, there is emerging evidence which suggests that mRNA expression patterns are necessary but are by themselves insufficient for the quantitative description of biological systems. This evidence includes discoveries of posttranscriptional mechanisms controlling the protein translation rate (15), the half-lives of specific proteins or mRNAs (33), and the intracellular location and molecular association of the protein products of expressed genes (32).

Proteome analysis, defined as the analysis of the protein complement expressed by a genome (26), has been suggested as an approach to the quantitative description of the state of a biological system by the quantitative analysis of protein expression profiles (36). Proteome analysis is conceptually attractive because of its potential to determine properties of biological systems that are not apparent by DNA or mRNA sequence analysis alone. Such properties include the quantity of protein expression, the subcellular location, the state of modification, and the association with ligands, as well as the rate of change with time of such properties. In contrast to the genomes of a number of microorganisms (for a review, see reference 11) and the transcriptome of *Saccharomyces cerevisiae* (35), which have been entirely determined, no proteome map has been completed to date.

The most common implementation of proteome analysis is the combination of two-dimensional gel electrophoresis (2DE)

(isoelectric focusing-sodium dodecyl sulfate [SDS]-polyacrylamide gel electrophoresis) for the separation and quantitation of proteins with analytical methods for their identification. 2DE permits the separation, visualization, and quantitation of thousands of proteins reproducibly on a single gel (18, 24). By itself, 2DE is strictly a descriptive technique. The combination of 2DE with protein analytical techniques has added the possibility of establishing the identities of separated proteins (1, 2) and thus, in combination with quantitative mRNA analysis, of correlating quantitative protein and mRNA expression measurements of selected genes.

The recent introduction of mass spectrometric protein analysis techniques has dramatically enhanced the throughput and sensitivity of protein identification to a level which now permits the large-scale analysis of proteins separated by 2DE. The techniques have reached a level of sensitivity that permits the identification of essentially any protein that is detectable in the gels by conventional protein staining (9, 29). Current protein analytical technology is based on the mass spectrometric generation of peptide fragment patterns that are idiosyncratic for the sequence of a protein. Protein identity is established by correlating such fragment patterns with sequence databases (10, 22, 37). Sophisticated computer software (8) has automated the entire process such that proteins are routinely identified with no human interpretation of peptide fragment patterns.

In this study, we have analyzed the mRNA and protein levels of a group of genes expressed in exponentially growing cells of the yeast *S. cerevisiae*. Protein expression levels were quantified by metabolic labeling of the yeast proteins to a steady state, followed by 2DE and liquid scintillation counting of the selected, separated protein species. Separated proteins were identified by in-gel tryptic digestion of spots with subsequent analysis by microspray liquid chromatography-tandem mass spectrometry (LC-MS/MS) and sequence database searching. The corresponding mRNA transcript levels were calculated from SAGE frequency tables (35).

This study, for the first time, explores a quantitative comparison of mRNA transcript and protein expression levels for a relatively large number of genes expressed in the same metabolic state. The resultant correlation is insufficient for predic-

* Corresponding author. Mailing address: Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195-7730. Phone: (206) 221-4196. Fax: (206) 685-7301. E-mail: ruedi@u.washington.edu.

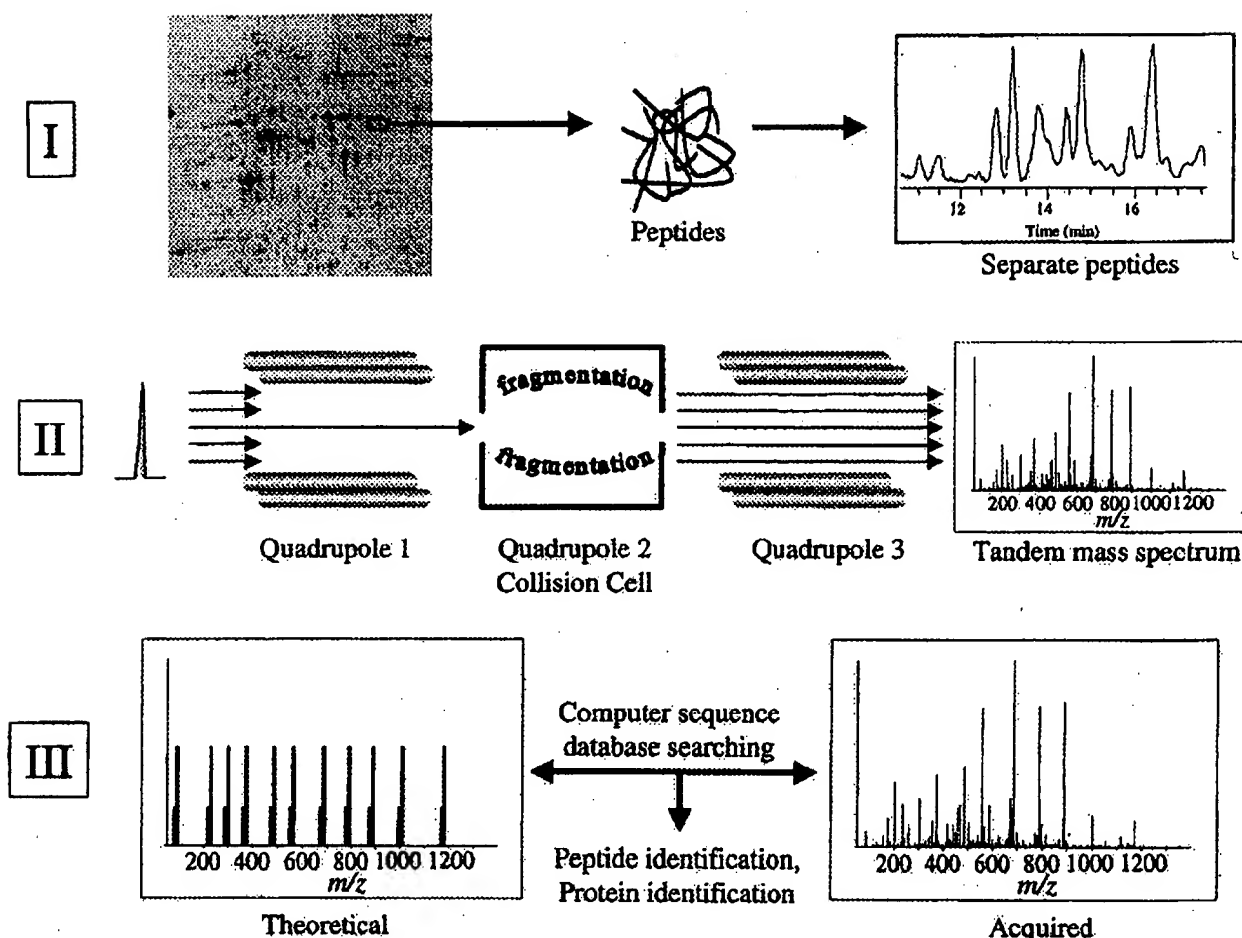


FIG. 1. Schematic illustration of proteome analysis by 2DE and mass spectrometry. In part I, proteins are separated by 2DE, stained spots are excised and subjected to in-gel digestion with trypsin, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. In part II, a peptide is shown eluting from the column in part I. The peptide is ionized by electrospray ionization and enters the mass spectrometer. The mass of the ionized peptide is detected, and the first quadrupole mass filter allows only the specific mass-to-charge ratio of the selected peptide ion to pass into the collision cell. In the collision cell, the energized, ionized peptides collide with neutral argon gas molecules. Fragmentation of the peptide is essentially random but occurs mainly at the peptide bonds, resulting in smaller peptides of differing lengths (masses). These peptide fragments are detected as a tandem mass (MS/MS) spectrum in the third quadrupole mass filter where two ion series are recorded simultaneously, one each from sequencing inward from the N and C termini of the peptide, respectively. In part III, the MS/MS spectrum from the selected, ionized peptide is compared to predicted tandem mass spectra computer generated from a sequence database. Provided that the peptide sequence exists in the database, the peptide and, by association, the protein from which the peptide was derived can be identified. Unambiguous protein identification is attained in a single analysis because multiple peptides are identified as being derived from the same protein.

tion of protein levels from mRNA transcript levels. We have also compared the relative amounts of protein and mRNA with the respective codon bias values for the corresponding genes. This comparison indicates that codon bias by itself is insufficient to accurately predict either the mRNA or the protein expression levels of a gene. In addition, the results demonstrate that only highly expressed proteins are detectable by 2DE separation of total cell lysates and that therefore the construction of complete proteome maps with current technology will be very challenging, irrespective of the type of organism.

MATERIALS AND METHODS

Yeast strain and growth conditions. The source of protein and message transcripts for all experiments was YPH499 (*MATa ura3-52 his2-801 ade2-101 leu2-Δ1 his3-Δ200 trp1-Δ63*) (30). Logarithmically growing cells were obtained by growing yeast cells to early log phase (3×10^6 cells/ml) in YPD rich medium (YPD supplemented with 6 mM uracil, 4.8 mM adenine, and 24 mM tryptophan) at 30°C (35). Metabolic labeling of protein was accomplished in YPD medium

exactly as described elsewhere (4) with the exception that 1 ml of cells was labeled with 3 mCi to offset methionine present in YPD medium. Protein was harvested as described by Garrels and coworkers (12). Harvested protein was lyophilized, resuspended in isoelectric focusing gel rehydration solution, and stored at -80°C.

2DE. Soluble proteins were run in the first dimension by using a commercial flatbed electrophoresis system (Multiphor II; Pharmacia Biotech). Immobilized polyacrylamide gel (IPG) dry strips with nonlinear pH 3.0 to 10.0 gradients (Amersham-Pharmacia Biotech) were used for the first-dimension separation. Forty micrograms of protein from whole-cell lysates was mixed with IPG strip rehydration buffer (8 M urea, 2% Nonidet P-40, 10 mM dithiothreitol), and 250 to 380 μ l of solution was added to individual lanes of an IPG strip rehydration tray (Amersham-Pharmacia Biotech). The strips were allowed to rehydrate at room temperature for 1 h. The samples were run at 300 V–10 mA–5 W for 2 h, then ramped to 3,500 V–10 mA–5 W over a period of 3 h, and then kept at 3,500 V–10 mA–5 W for 15 to 19 h. At the end of the first-dimension run (60 to 70 kV·h), the IPG strips were reequilibrated for 8 min in 2% (wt/vol) dithiothreitol in 2% (wt/vol) SDS–6 M urea–30% (wt/vol) glycerol–0.05 M Tris HCl (pH 6.8) and for 4 min in 2.5% iodoacetamide in 2% (wt/vol) SDS–6 M urea–30% (wt/vol) glycerol–0.05 M Tris HCl (pH 6.8). Following reequilibration, the strips were transferred and apposed to 10% polyacrylamide second-dimension gels. Polyacrylamide gels were poured in a casting stand with 10% acrylamide–2.67% piperazine diacrylamide–0.375 M Tris base–HCl (pH 8.8)–0.1% (wt/vol) SDS–0.05%

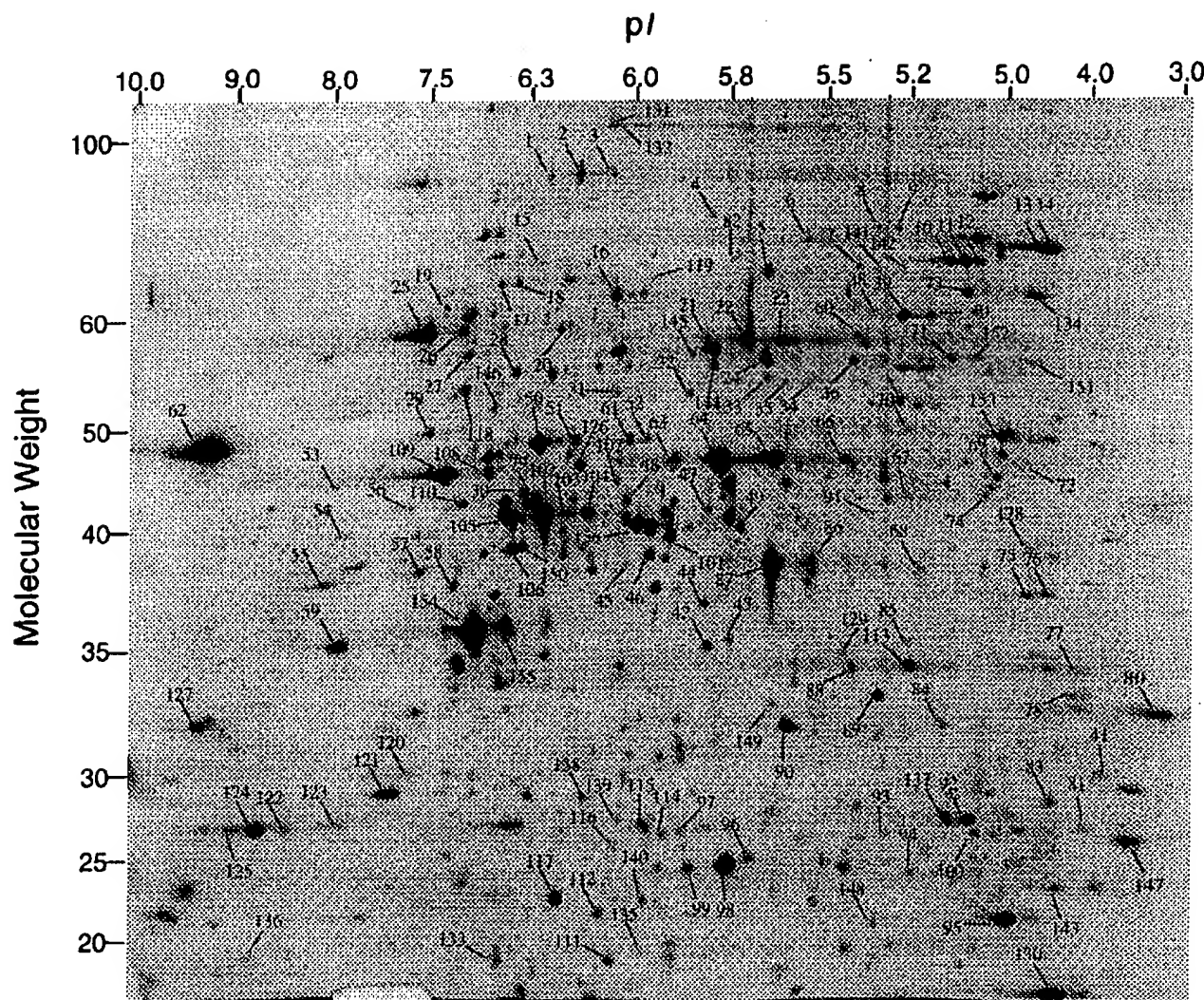


FIG. 2. 2D silver-stained gel of the proteins in yeast total cell lysate. Proteins were separated in the first dimension (horizontal) by isoelectric focusing and then in the second dimension (vertical) by molecular weight sieving. Protein spots (156) were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. Spots were excised, and the corresponding protein was identified by mass spectrometry and database searching. The spots are labeled on the gel and correspond to the data presented in Table 1. Molecular weights are given in thousands.

(wt/vol) ammonium persulfate–0.05% TEMED (N,N,N',N' -tetramethylethylenediamine) in Milli-Q water. The apparatus used to run second-dimension gels was a noncommercial apparatus from Oxford Glycosciences, Inc. Once the IPG strips were apposed to the second-dimension gels, they were immediately run at 50 mA (constant)–500 V–85 W for 20 min, followed by 200 mA (constant)–500 V–85 W until the buffer front line was 10 to 15 mm from the bottom of the gel. Gels were removed and silver stained according to the procedure of Shevchenko et al. (29).

Protein identification. Gels were exposed to X-ray film overnight, and then the silver staining and film were used to excise 156 spots of varying intensities, molecular weights, and isoelectric focusing points. In order to increase the detection limit by mass spectrometry, spots were cut out and pooled from up to four identical cold; silver-stained gels. In-gel tryptic digests of pooled spots were performed as described previously (29). Tryptic peptides were analyzed by microcapillary LC-MS with automated switching to MS/MS mode for peptide fragmentation. Spectra were searched against the composite OWL protein sequence database (version 30.2; 250,514 protein sequences) (24a) by using the computer program Sequest (8), which matches theoretical and acquired tandem mass spectra. A protein match was determined by comparing the number of peptides identified and their respective cross-correlation scores. All protein identifications were verified by comparison with theoretical molecular weights and isoelectric points.

mRNA quantitation. Velculescu and coworkers have previously generated frequency tables for yeast mRNA transcripts from the same strain grown under the same stated conditions as described herein (35). The SAGE technology is based on two main principles. First, a short sequence tag (15 bp) that contains sufficient information uniquely to identify a transcript is generated. A single tag is usually generated from each mRNA transcript in the cell which corresponds to 15 bp at the 3'-most cutting site for *NotI*. Second, many transcript tags can be concatenated into a single molecule and then sequenced, revealing the identity of multiple tags simultaneously. Over 20,000 transcripts were sequenced from yeast strain YPH499 growing at mid-log phase on glucose. Assuming the previously derived estimate of 15,000 mRNA molecules per cell (16), this would represent a 1.3-fold coverage even for mRNA molecules present at a single copy per cell and would provide a 72% probability of detecting such transcripts. Computer software which took for input the gene detected, examined the nucleotide sequence, and performed the calculation as described by Velculescu and coworkers (35) was written. In practice, we found that for 21 of 128 (16%) genes examined viable mRNA levels from SAGE data could not be calculated. This was because (i) no CATG site was found in the open reading frame (ORF), (ii) a CATG site was found but the corresponding 10-bp putative SAGE tag was not found in the frequency tables, or (iii) identical putative SAGE tags were present for multiple genes (e.g., TDH2_YEAST and TDH3_YEAST).

TABLE 1. Expressed genes identified from 2D gel in Fig. 2

Mol wt	pI	Spot no.	YPD gene name ^a	Protein abundance (10 ³ copies/cell)	mRNA abundance (copies/cell)	Codon bias
17,259	6.75	133	CPR1	15.2	61.7	0.769
18,702	4.80	83	EGD2	20.1	5.2	0.724
18,726	4.44	147	YKL056C	61.2	88.4	0.831
18,978	5.95	135	YER067W	3.7	6.7	0.118
19,108	5.04	130	YLR109W	94.4	9.7	0.680
19,681	9.08	136	ATP7	11.0	NA ^{b,c}	0.246
20,505	6.07	111	GUK1	16.5	3.7	0.422
21,444	5.25	148	SAR1	5.4	10.4	0.455
21,583	4.98	95	TSA1	110.6	40.1	0.845
22,602	4.30	80	EFB1	66.1	23.8	0.875
23,079	6.29	112	SOD2	12.6	2.2	0.351
23,743	5.44	137	HSP26	NA ^d	0.7	0.434
24,033	5.97	96	ADK1	17.4	16.4	0.656
24,058	4.43	143	YKL117W	29.2	10.4	0.339
24,353	6.30	140	TFS1	8.1	0.7	0.146
24,662	5.85	99	URA5	25.4	6.0	0.359
24,808	6.33	97	GSP1	26.3	5.2	0.735
24,908	8.73	122	RPS5	18.6	NA ^c	0.899
25,081	4.65	81	MRP8	9.3	NA ^c	0.241
25,960	6.06	116	RPE1	5.8	0.7	0.372
26,378	9.55	127	RPS3	96.8	NA ^c	0.863
26,467	5.18	100	VMA4	10.5	3.7	0.427
26,661	5.84	98	TPI1	NA ^d	NA ^c	0.900
27,156	5.56	93	PRE8	6.9	0.7	0.129
27,334	6.13	115	YHR049W	18.4	2.2	0.520
27,472	5.33	92	YNL010W	31.6	3.7	0.421
27,480	8.95	123	GPM1	10.0	169.4	0.902
27,480	8.95	124	GPM1	231.4	169.4	0.902
27,480	8.95	125	GPM1	7.5	169.4	0.902
27,809	5.97	139	HOR2	5.7	0.7	0.381
27,874	4.46	78	YST1	13.6	52.8	0.805
28,595	4.51	41	PUP2	4.4	0.7	0.147
29,156	6.59	114	YMR226C	14.5	2.2	0.283
29,244	8.40	120	DPM1	5.0	11.2	0.362
29,443	5.91	48	PRE4	3.4	3.7	0.162
30,012	6.39	138	PRB1	21.2	1.5	0.449
30,073	4.63	77	BMH1	14.7	28.2	0.454
30,296	7.94	121	OMP2	67.4	41.6	0.499
30,435	6.34	89	GPP1	70.2	11.2	0.703
31,332	5.57	88	ILV6	13.9	3.0	0.402
32,159	5.46	113	IPP1	63.1	3.7	0.752
32,263	6.00	149	HIS1	22.4	4.5	0.232
33,311	5.35	84	SPE3	15.1	6.7	0.468
34,465	5.60	129	ADE1	8.7	5.2	0.305
34,762	5.32	85	SEC14	10.9	6.0	0.373
34,797	5.85	42	URA1	49.5	8.9	0.237
34,799	6.04	90	BEL1	103.2	81.0	0.875
35,556	5.97	43	YDL124W	6.4	4.5	0.206
35,619	8.41	59	TDH1	69.8	32.7 ^c	0.940
35,650	5.49	68	CAR1	5.2	3.0	0.339
35,712	6.72	117	TDH2	49.6	473.0 ^c	0.982
35,712	6.72	154	TDH2	863.5	473.0 ^c	0.982
35,712	6.72	155	TDH2	79.4	473.0 ^c	0.982
36,272	4.85	128	APA1	8.7	0.7	0.425
36,358	5.05	75	YJR105W	17.6	17.1	0.522
36,358	5.05	76	YJR105W	27.5	17.1	0.522
36,596	6.37	79	ADH2	58.9	260.0 ^c	0.711
36,714	6.30	102	ADH1	746.1	260.0	0.913
36,714	6.30	103	ADH1	17.6	260.0	0.913
36,714	6.30	104	ADH1	61.4	260.0	0.913
36,714	6.30	105	ADH1	52.7	260.0	0.913
37,033	6.23	44	TAL1	44.8	3.7	0.701
37,796	7.36	57	IDH2	29.4	6.7	0.330
37,886	6.49	106	ILV5	76.0	4.5	0.892
38,700	7.83	55	BAT1	30.9	11.2	0.469
38,702	6.24	46	QCR2	NA ^d	2.2	0.326

Continued

TABLE 1—Continued

Mol wt	pI	Spot no.	YPD gene name ^a	Protein abundance (10 ³ copies/cell)	mRNA abundance (copies/cell)	Codon bias
39,477	5.58	86	FBA1	17.8	183.6	0.935
39,477	5.58	87	FBA1	427.2	183.6	0.935
39,540	6.50	150	HOM2	60.3	4.5	0.592
39,561	6.12	156	PSA1	96.4	27.5	0.718
41,158	6.01	49	YNL134C	14.9	1.5	0.316
41,623	7.18	58	BAT2	19.0	8.9	0.250
41,728	7.29	110	ERG10	24.1	4.5	0.543
41,900	5.42	74	TOM40	22.3	2.2	0.375
42,402	6.29	45	CYS3	6.7	8.9	0.621
42,883	5.63	67	DYS1	15.8	5.2	0.526
43,409	6.31	107	SER1	10.5	1.5	0.292
43,421	5.59	91	ERG6	2.2	14.1	0.408
44,174	7.32	56	YBR025C	13.1	6.0	0.684
44,682	4.99	72	TIF1	2.9	39.4	0.834
44,707	7.77	108	PGK1	23.7	165.7	0.897
44,707	7.77	109	PGK1	315.2	165.7	0.897
46,080	6.72	30	CAR2	15.4	NA ^c	0.495
46,383	8.52	53	IDP1	7.7	0.7	0.436
46,553	5.98	47	IDP2	32.4	NA ^c	0.197
46,679	6.39	50	ENO1	35.4	0.7	0.930
46,679	6.39	51	ENO1	6.6	0.7	0.930
46,679	6.39	52	ENO1	2.2	0.7	0.930
46,773	5.82	63	ENO2	15.5	289.1	0.960
46,773	5.82	64	ENO2	635.5	289.1	0.960
46,773	5.82	65	ENO2	93.0	289.1	0.960
46,773	5.82	66	ENO2	31.0	289.1	0.960
47,402	6.09	126	COR1	2.5	0.7	0.422
47,666	8.98	54	AAT2	11.7	6.0	0.338
48,364	5.25	73	WTM1	74.5	13.4	0.365
48,530	6.20	61	MET17	38.1	29.0	0.576
48,904	5.18	69	LYS9	16.2	3.7	0.463
48,987	4.90	153	SUP45	29.6	11.9	0.377
49,727	5.47	70	PRO2	13.6	5.2	0.297
49,912	9.27	62	TEF2	558.5	282.0	0.932
50,444	5.67	35	YDR190C	4.8	2.2	0.228
50,837	6.11	32	YEL047C	3.8	1.5	0.387
50,891	4.59	151	TUB2	11.2	7.4	0.404
51,547	6.80	27	LPD1	18.9	2.2	0.351
52,216	7.25	29	SHM2	19.7	7.4	0.722
52,859	5.54	37	YFR044C	30.2	6.7	0.442
53,798	5.19	71	HXK2	26.5	7.4	0.756
53,803	6.05	145	GYP6	4.4	0.7	0.147
54,403	5.29	39	ALD6	37.7	2.2	0.664
54,403	5.29	40	ALD6	6.6	2.2	0.664
54,502	6.20	31	ADE13	6.3	1.5	0.417
54,543	7.75	25	PYK1	225.3	101.8	0.965
54,543	7.75	26	PYK1	39.8	101.8	0.965
55,221	6.66	146	YEL071W	16.3	3.0	0.244
55,295	4.35	134	PDH1	66.2	14.1	0.589
55,364	5.98	24	GLK1	22.6	6.0	0.237
55,481	7.97	118	ATP1	21.6	2.2	0.637
55,886	6.47	28	CYS4	22.2	NA ^c	0.444
56,167	5.83	33	ARO8	14.3	3.0	0.324
56,167	5.83	34	ARO8	9.1	3.0	0.324
56,584	6.36	20	CYB2	18.9	NA ^c	0.259
57,366	5.53	60	FKS2	2.3	0.7	0.451
57,383	5.98	144	ZWF1	5.6	0.7	0.215
57,464	5.49	36	TEH4	21.4	3.7	0.508
57,512	5.50	7	SRV2	6.5	NA ^c	0.260
57,727	4.92	152	VMA2	33.7	8.9	0.546
58,573	6.47	17	ACH1	4.4	1.5	0.327
58,573	6.47	18	ACH1	5.4	1.5	0.327
61,353	5.87	21	PDC1	6.5	200.7	0.962
61,353	5.87	22	PDC1	303.2	200.7	0.962
61,353	5.87	23	PDC1	16.3	200.7	0.962
61,649	5.54	38	CCT8	2.2	1.5	0.271

Continued on following page

TABLE 1—Continued

Mol wt	pI	Spot no.	YPD gene name ^a	Protein abundance (10 ³ copies/cell)	mRNA abundance (copies/cell)	Codon bias
61,902	6.21	101	PDC5	4.3	NA ^c	0.828
62,266	6.19	16	ICL1	20.1	NA ^c	0.327
62,862	8.02	19	ILV3	5.3	4.5	0.548
63,082	6.40	119	PGM2	2.2	3.0	0.402
64,335	5.77	5	PAB1	30.4	1.5	0.616
66,120	5.42	8	STI1	6.7	0.7	0.313
66,120	5.42	9	STI1	6.4	0.7	0.313
66,450	5.29	141	SSB2	7.0	NA ^c	0.880
66,450	5.29	142	SSB2	2.3	NA ^c	0.880
66,456	5.23	10	SSB1	64.5	79.5	0.907
66,456	5.23	11	SSB1	59.0	79.5	0.907
66,456	5.23	12	SSB1	13.7	79.5	0.907
68,397	5.82	82	LEU4	3.1	3.0	0.407
69,313	4.90	13	SSA2	24.3	18.6	0.892
69,313	4.90	14	SSA2	77.1	18.6	0.892
74,378	8.46	15	YKL029C	2.8	3.7	0.353
75,396	5.82	6	GRS1	5.5	7.4	0.500
85,720	6.25	1	MET6	2.0	NA ^c	0.772
85,720	6.25	2	MET6	10.9	NA ^c	0.772
85,720	6.25	3	MET6	1.4	NA ^c	0.772
93,276	6.11	131	EFT1	17.9	41.6	0.890
93,276	6.11	132	EFT1	5.7	41.6	0.890
102,064 ^e	6.61 ^e	94	ADE3	4.8	5.2	0.423
107,482 ^e	5.33 ^e	4	MCM3	2.7	NA ^c	0.240

^a YPD gene names are available from the YPD website (39).^b NA, calculation could not be performed or was not available.^c mRNA data inconclusive or NA.^d No methionines in predicted ORF; therefore, protein concentration was not determined.^e Measured molecular weight or pI did not match theoretical molecular weight or pI.

Protein quantitation. [³⁵S]methionine-labeled gels were exposed to X-ray film overnight, and then the silver stain and film were used to excise 156 spots of varying intensities, molecular weights, and pIs. The excised spots were placed in 0.6-ml microcentrifuge tubes, and scintillation cocktail (100 μ l) was added. The samples were vortexed and counted. In addition, two parallel gels were electroblotted to polyvinylidene difluoride membranes. The membranes were exposed to X-ray film, and four intense single spots were excised from each membrane and subjected to amino acid analysis. For these four spots, a mean of 209 ± 4 cpm/pmol of protein/methionine was found. This number was used to quantitate all remaining spots in conjunction with the number of methionines present in the protein.

To ensure that proteins were labeled to equilibrium, parallel 2D gels were prepared and run on yeast metabolically labeled for 1, 2, 6, or 18 h. The corresponding 156 spots were excised from each gel, and radioactivity was measured by liquid scintillation counting for each spot. Calculated protein levels were highly reproducible for all time points measured after 1 h.

Calculation of codon bias and predicted half-life. Codon bias values were extracted from the YPD spreadsheet (17). Protein half-lives were calculated based on the N-end rule (33). When the N-terminal processing was not known experimentally, it was predicted based on the affinity of methionine aminopeptidase (31).

RESULTS

Characteristics of proteome approach. Nearly every facet of proteome analysis hinges on the unambiguous identification of large numbers of expressed proteins in cells. Several techniques have been described previously for the identification of proteins separated by 2DE, including N-terminal and internal sequencing (1, 2), amino acid analysis (38), and more recently mass spectrometry (25). We utilized techniques based on mass spectrometry because they afford the highest levels of sensitivity and provide unambiguous identification. The specific procedure used is schematically illustrated in Fig. 1 and is based on three principles. First, proteins are removed from the gel by

proteolytic in-gel digestion, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. Second, the eluting peptides are ionized and detected, and the specific peptide ions are selected and fragmented by the mass spectrometer. To achieve this, the mass spectrometer switches between the MS mode (for peptide mass identification) and the MS/MS mode (for peptide characterization and sequencing). Selected peptides are fragmented by a process called collision-induced dissociation (CID) to generate a tandem mass spectrum (MS/MS spectrum) that contains the peptide sequence information. Third, individual CID mass spectra are then compared by computer algorithms to predicted spectra from a sequence database. This results in the identification of the peptide and, by association, the protein(s) in the spot. Unambiguous protein identification is attained in a single analysis by the detection of multiple peptides derived from the same protein.

Protein identification. Yeast total cell protein lysate (40 μ g), metabolically labeled with [³⁵S]methionine, was electrophoretically separated by isoelectric focusing in the first dimension and by SDS-10% polyacrylamide gel electrophoresis in the second dimension. Proteins were visualized by silver staining and by autoradiography. Of the more than 1,000 proteins visible by silver staining, 156 spots were excised from the gel and subjected to in-gel tryptic digestion, and the resulting peptides were analyzed and identified by microspray LC-MS/MS techniques as described above. The proteins in this study were all identified automatically by computer software with no human interpretation of mass spectra. They are indicated in Fig. 2 and detailed in Table 1.

The CID spectra shown in Fig. 3 indicate that the quality of the identification data generated was suitable for unambiguous protein identification. The spectra represent the amino acid sequences of tryptic peptides NSGDIVNLGSIAGR (Fig. 3A) and FAVGAFTDSLRL (Fig. 3B). Both peptides were derived from protein S57593 (hypothetical protein YMR226C), which migrated to spot 114 (molecular weight, 29,156; pI, 6.59) in the 2D gel in Fig. 2. Five other peptides from the same analysis were also computer matched to the same protein sequence.

Protein and mRNA quantitation. For the 156 genes investigated, the protein expression levels ranged from 2,200 (PGM2) to 863,000 (TDH2/TDH3) copies/cell. The levels of mRNA for each of the genes identified were calculated from SAGE frequency tables (35). These tables contain the mRNA levels for 4,665 genes in yeast strain YPH499 grown to mid-log phase in YPD medium on glucose as a carbon source. In some instances, the mRNA levels could not be calculated for reasons stated in Materials and Methods. For the proteins analyzed in this study, mean transcript levels varied from 0.7 to 473 copies/cell.

Selection of the sample population for mRNA-protein expression level correlation. The protein spots selected for identification were selected from spots visible by silver staining in the 2D gel. An attempt was made not to include spots where overlap with other spots was readily apparent. The number of proteins identified was 156 (Table 1). Some proteins migrated to more than one spot (presumably due to differential protein processing or modifications), and protein levels from these spots were calculated by integrating the intensities of the different spots. The 156 protein spots analyzed represented the products of 128 different genes. Genes were excluded from the correlation analysis only if part of the data set was missing; i.e., genes were excluded if (i) no mRNA expression data were available for the protein or putative SAGE tags were ambiguous, (ii) the amino acid sequence did not contain methionine, (iii) more than a single protein was conclusively identified as

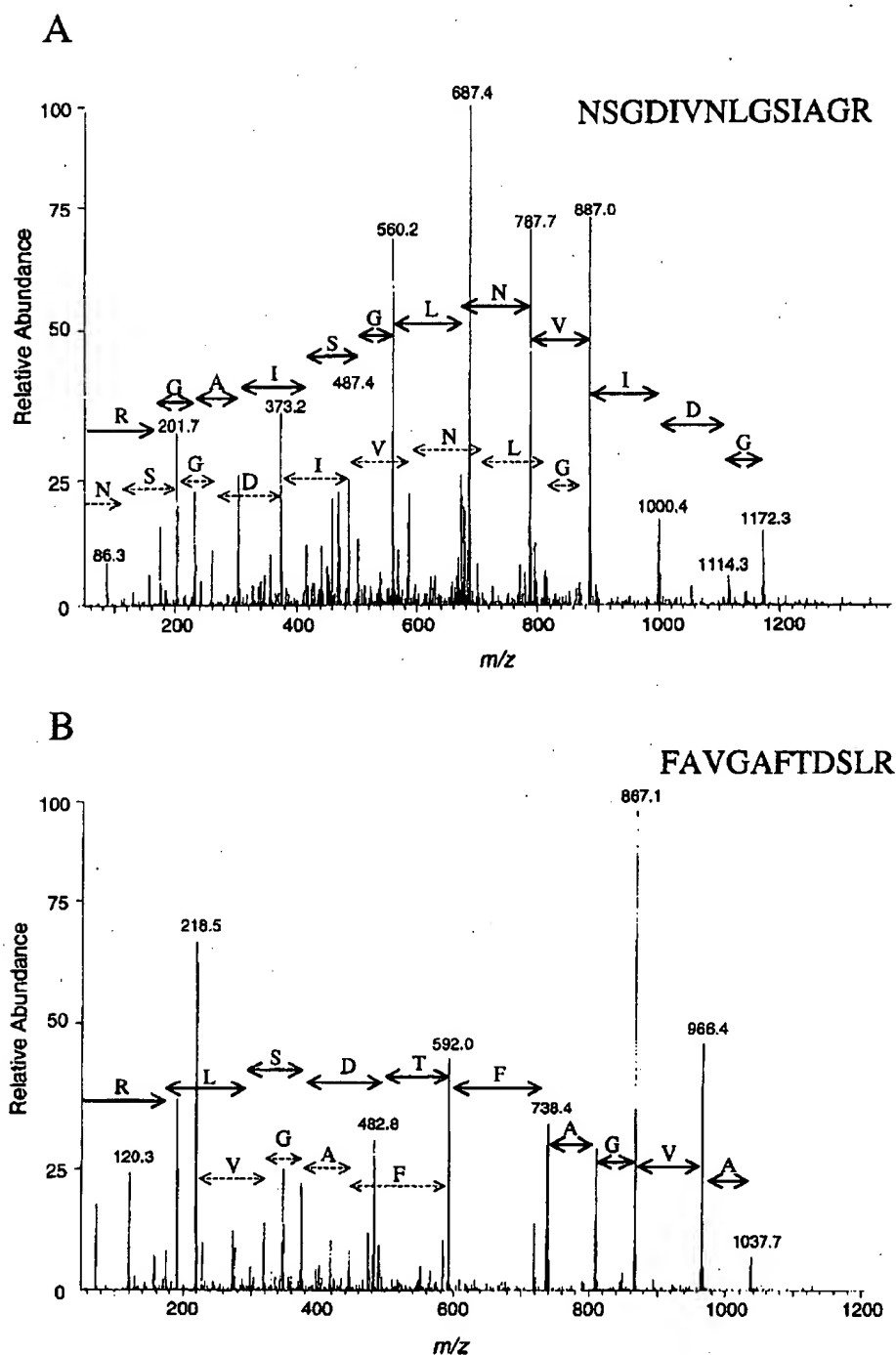


FIG. 3. Tandem mass (MS/MS) spectra resulting from analysis of a single spot on a 2D gel. The first quadrupole selected a single mass-to-charge ratio (m/z) of 687.2 (A) or 592.6 (B), while the collision cell was filled with argon gas, and a voltage which caused the peptide to undergo fragmentation by CID was applied. The third quadrupole scanned the mass range from 50 to 1,400 m/z . The computer program Sequest (8) was utilized to match MS/MS spectra to amino acid sequence by database searching. Both spectra matched peptides from the same protein, S57593 (yeast hypothetical protein YMR226C). Five other peptides from the same analysis were matched to the same protein.

migrating to the same gel spot, or (iv) the theoretical and observed pIs and molecular weights could not be reconciled. After these criteria were applied, the number of genes used in the correlation analysis was 106.

Codon bias and predicted half-lives. Codon bias is thought to be an indicator of protein expression, with highly expressed proteins having large codon bias values. The codon bias distribution for the entire set of more than 6,000 predicted yeast

gene ORFs is presented in Fig. 4A. The interval with the largest frequency of genes is between the codon bias values of 0.0 and 0.1. This segment contains more than 2,500 genes. The distribution of the codon bias values of the 128 different genes found in this study (all protein spots from Fig. 2) is shown in Fig. 4B, and protein half-lives (predicted from applying the N-end rule [33] to the experimentally determined or predicted protein N termini) are shown in Fig. 4C. No genes were identified with codon bias values less than 0.1 even though thousands of genes exist in this category. In addition, nearly all of the proteins identified had long predicted half-lives (greater than 30 h).

Correlation of mRNA and protein expression levels. The correlation between mRNA and protein levels of the genes selected as described above is shown in Fig. 5. For the entire group (106 genes) for which a complete data set was generated, there was a general trend of increased protein levels resulting from increased mRNA levels. The Pearson product moment correlation coefficient for the whole data set (106 genes) was 0.935. This number is highly biased by a small number of genes with very large protein and message levels. A more representative subset of the data is shown in the inset of Fig. 5. It shows genes for which the message level was below 10 copies/cell and includes 69% (73 of 106 genes) of the data used in the study. The Pearson product moment correlation coefficient for this data set was only 0.356. We also found that levels of protein expression coded for by mRNA with comparable abundance varied by as much as 30-fold and that the mRNA levels coding for proteins with comparable expression levels varied by as much as 20-fold.

The distortion of the correlation value induced by the uneven distribution of the data points along the *x* axis is further demonstrated by the analysis in Fig. 6. The 106 samples included in the study were ranked by protein abundance, and the Pearson product moment correlation coefficient was repeatedly calculated after including progressively more, and higher-abundance, proteins in each calculation. The correlation values remained relatively stable in the range of 0.1 to 0.4 if the lowest-expressed 40 to 95 proteins used in this study were included. However, the correlation value steadily climbed by the inclusion of each of the 11 very highly expressed proteins.

Correlation of protein and mRNA expression levels with codon bias. Codon bias is the propensity for a gene to utilize the same codon to encode an amino acid even though other codons would insert the identical amino acid in the growing polypeptide sequence. It is further thought that highly expressed proteins have large codon biases (3). To assess the value of codon bias for predicting mRNA and protein levels in exponentially growing yeast cells, we plotted the two experimental sets of data versus the codon bias (Fig. 7). The distribution patterns for both mRNA and protein levels with respect to codon bias were highly similar. There was high variability in the data within the codon bias range of 0.8 to 1.0. Although a large codon bias generally resulted in higher protein and message expression levels, codon bias did not appear to be predictive of either protein levels or mRNA levels in the cell.

DISCUSSION

The desired end point for the description of a biological system is not the analysis of mRNA transcript levels alone but also the accurate measurement of protein expression levels and their respective activities. Quantitative analysis of global mRNA levels currently is a preferred method for the analysis of the state of cells and tissues (11). Several methods which either provide absolute mRNA abundance (34, 35) or relative

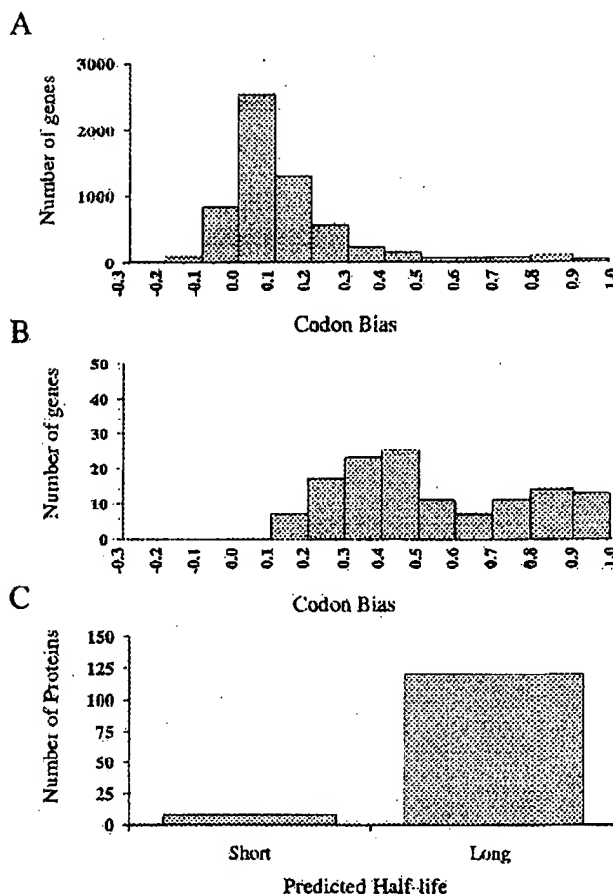


FIG. 4. Current proteome analysis technology utilizing 2DE without pre-enrichment samples mainly highly expressed and long-lived proteins. Genes encoding highly expressed proteins generally have large codon bias values. (A) Distribution of the yeast genome (more than 6,000 genes) based on codon bias. The interval with the largest frequency of genes is 0.0 to 0.1, with more than 2,500 genes. (B) Distribution of the genes from identified proteins in this study based on codon bias. No genes with codon bias values less than 0.1 were detected in this study. (C) Distribution of identified proteins in this study based on predicted half-life (estimated by N-end rule).

mRNA levels in comparative analyses (20, 27) have been described elsewhere. The techniques are fast and exquisitely sensitive and can provide mRNA abundance for potentially any expressed gene. Measured mRNA levels are often implicitly or explicitly extrapolated to indicate the levels of activity of the corresponding protein in the cell. Quantitative analysis of protein expression levels (proteome analysis) is much more time-consuming because proteins are analyzed sequentially one by one and is not general because analyses are limited to the relatively highly expressed proteins. Proteome analysis does, however, provide types of data that are of critical importance for the description of the state of a biological system and that are not readily apparent from the sequence and the level of expression of the mRNA transcript. This study attempts to examine the relationship between mRNA and protein expression levels for a large number of expressed genes in cells representing the same state.

Limits in the sensitivity of current protein analysis technology precluded a completely random sampling of yeast proteins. We therefore based the study on those proteins visible by silver

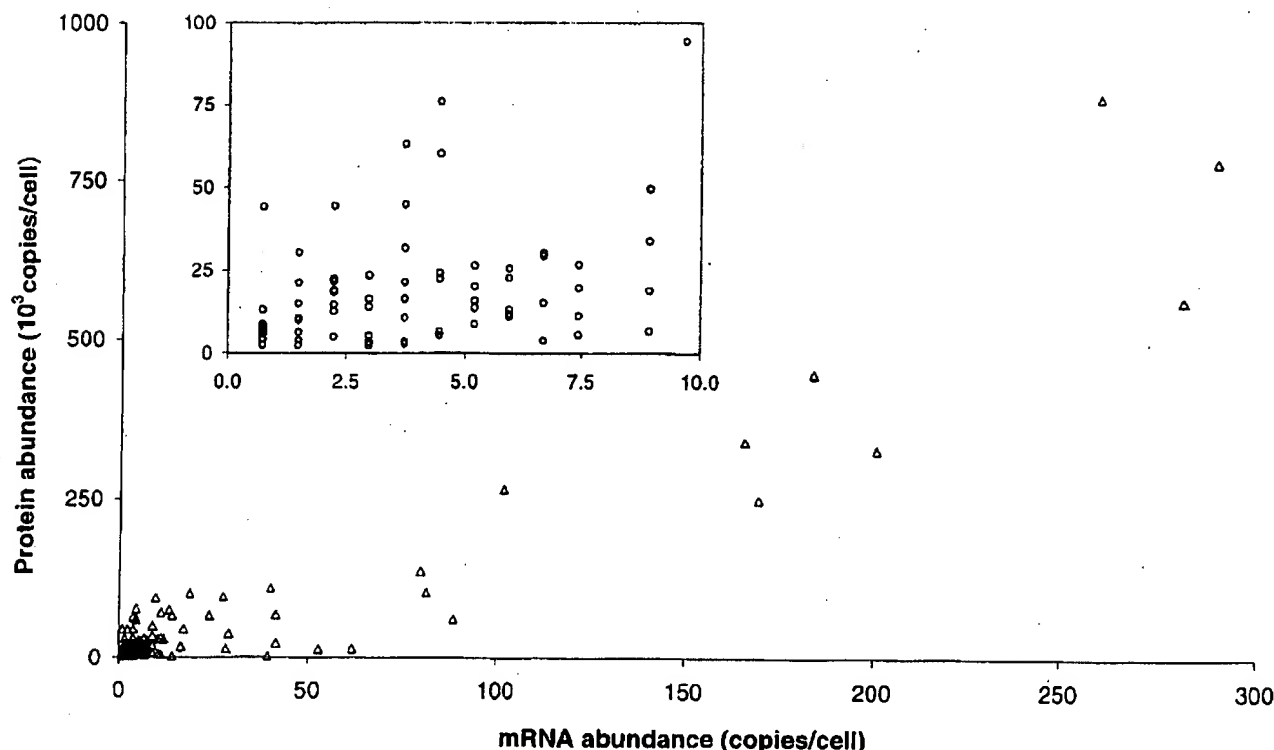


FIG. 5. Correlation between protein and mRNA levels for 106 genes in yeast growing at log phase with glucose as a carbon source. mRNA and protein levels were calculated as described in Materials and Methods. The data represent a population of genes with protein expression levels visible by silver staining on a 2D gel chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. The inset shows the low-end portion of the main figure. It contains 69% of the original data set. The Pearson product moment correlation for the entire data set was 0.935. The correlation for the inset containing 73 proteins (69%) was only 0.356.

staining on a 2D gel. Of the more than 1,000 visible spots, 156 were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities displayed on the 2D protein pattern. The genes identified in this study shared a number of properties. First, all of the proteins in this study had a codon bias of greater than 0.1 and 93% were greater than 0.2 (Fig. 4B). Second, with few exceptions, the proteins in this study had long predicted half-lives according to the N-end rule (Fig. 4C). Third, low-abundance proteins with regulatory functions such as transcription factors or protein kinases were not identified.

Because the population of proteins used in this study appears to be fairly homogeneous with respect to predicted half-life and codon bias, it might be expected that the correlation of the mRNA and protein expression levels would be stronger for this population than for a random sample of yeast proteins. We tested this assumption by evaluating the correlation value if different subsets of the available data were included in the calculation. The 106 proteins were ranked from lowest to highest protein expression level, and the trend in the correlation value was evaluated by progressively including more of the higher-abundance proteins in the calculation (Fig. 6). The correlation value when only the lower-abundance 40 to 93 proteins were examined was consistently between 0.1 and 0.4. If the 11 most abundant proteins were included, the correlation steadily increased to 0.94. We therefore expect that the correlation for all yeast proteins or for a random selection would be less than 0.4. The observed level of correlation between mRNA and protein expression levels suggests the importance

of posttranslational mechanisms controlling gene expression. Such mechanisms include translational control (15) and control of protein half-life (33). Since these mechanisms are also active in higher eukaryotic cells, we speculate that there is no predictive correlation between steady-state levels of mRNA and those of protein in mammalian cells.

Like other large-scale analyses, the present study has several potential sources of error related to the methods used to determine mRNA and protein expression levels. The mRNA levels were calculated from frequency tables of SAGE data. This method is highly quantitative because it is based on actual sequencing of unique tags from each gene, and the number of times that a tag is represented is proportional to the number of mRNA molecules for a specific gene. This method has some limitations including the following: (i) the magnitude of the error in the measurement of mRNA levels is inversely proportional to the mRNA levels, (ii) SAGE tags from highly similar genes may not be distinguished and therefore are summed, (iii) some SAGE tags are from sequences in the 3' untranslated region of the transcript, (iv) incomplete cleavage at the SAGE tag site by the restriction enzyme can result in two tags representing one mRNA, and (v) some transcripts actually do not generate a SAGE tag (34, 35).

For the SAGE method, the error associated with a value increases with a decreasing number of transcripts per cell. The conclusions drawn from this study are dependent on the quality of the mRNA levels from previously published data (35). Since more than 65% of the mRNA levels included in this study were calculated to 10 copies/cell or less (40% were less

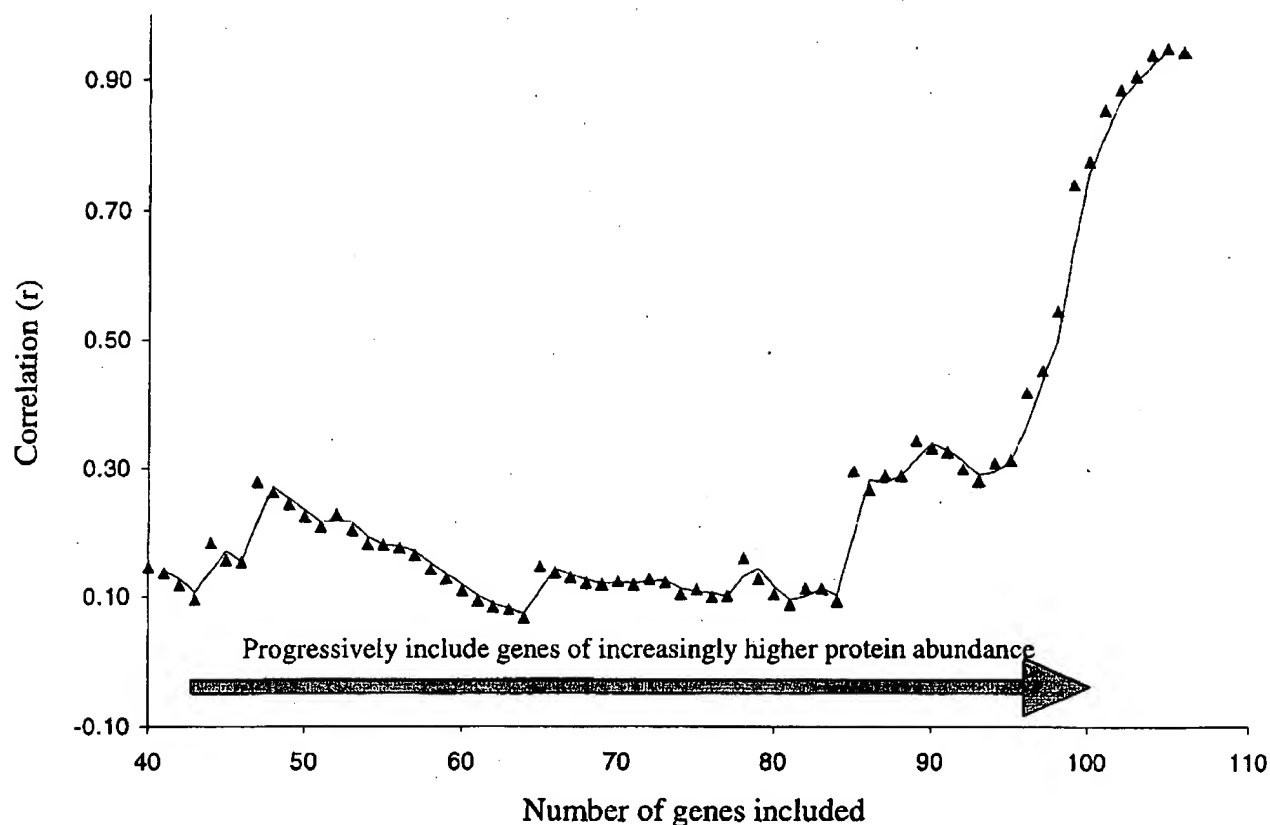


FIG. 6. Effect of highly abundant proteins on Pearson product moment correlation coefficient for mRNA and protein abundance in yeast. The set of 106 genes was ranked according to protein abundance, and the correlation value was calculated by including the 40 lowest-abundance genes and then progressively including the remaining 66 genes in order of abundance. The correlation value climbs as the final 11 highly abundant proteins are included.

than 4 copies/cell), the error associated with these values may be quite large. The mRNA levels were calculated from more than 20,000 transcripts. Assuming that the estimate of 15,000 mRNA molecules per cell is correct (16), this would mean that mRNA transcripts present at only a single copy per cell would be detected 72% of the time (35). The mRNA levels for each gene were carefully scrutinized, and only mRNA levels for which a high degree of confidence existed were included in the correlation value.

Protein abundance was determined by metabolic radiolabeling with [35 S]methionine. The calculation required knowledge of three variables: the number of methionines in the mature protein, the radioactivity contained in the protein, and the specific activity of the radiolabel normalized per methionine. The number of methionines per protein was determined from the amino acid sequence of the proteins identified by tandem mass spectrometry. For some proteins, it was not known whether the methionine of the nascent polypeptide was processed away. The N termini of those proteins were predicted based on the specificity of methionine aminopeptidase (31). If the N-terminal processing did not conform to the predicted specificity of processing enzymes, the calculation of the number of methionines would be affected. This discrepancy would affect most the quantitation of a protein with a very low number of methionines. The average number of calculated methionines per protein in this study was 7.2. We therefore expect the potential for erroneous protein quantitation due to unusual N-terminal processing to be small.

The amount of radioactivity contained in a single spot might be the sum of the radioactivity of comigrating proteins. Because protein identification was based on tandem mass spectrometric techniques, comigrating proteins could be identified. However, comigrating proteins were rarely detected in this study, most likely because relatively small amounts of total protein (40 μ g) were initially loaded onto the gels, which resulted in highly focused spots containing generally 1 to 25 ng of protein. Because of the relatively small amount loaded, the concentrations of any potentially comigrating protein would likely be below the limit of detection of the mass spectrometry technique used in this study (1 to 5 ng) and below the limit of visualization by silver staining (1 to 5 ng). In the overwhelming majority of the samples analyzed, numerous peptides from a single protein were detected. It is assumed that any comigrating proteins were at levels too low to be detected and that their influence in the calculation would be small.

The specific activity of the radiolabel was determined by relating the precise amount of protein present in selected spots of a parallel gel, as determined by quantitative amino acid composition analysis, to the number of methionines present in the sequence of those proteins and the radioactivity determined by liquid scintillation counting. It is possible that the resulting number might be influenced by unavoidable losses inherent in the amino acid analysis procedure applied. Because four different proteins were utilized in the calculation and the experiment was done in duplicate, the specific activity calculated is thought to be highly accurate. Indeed, the specific

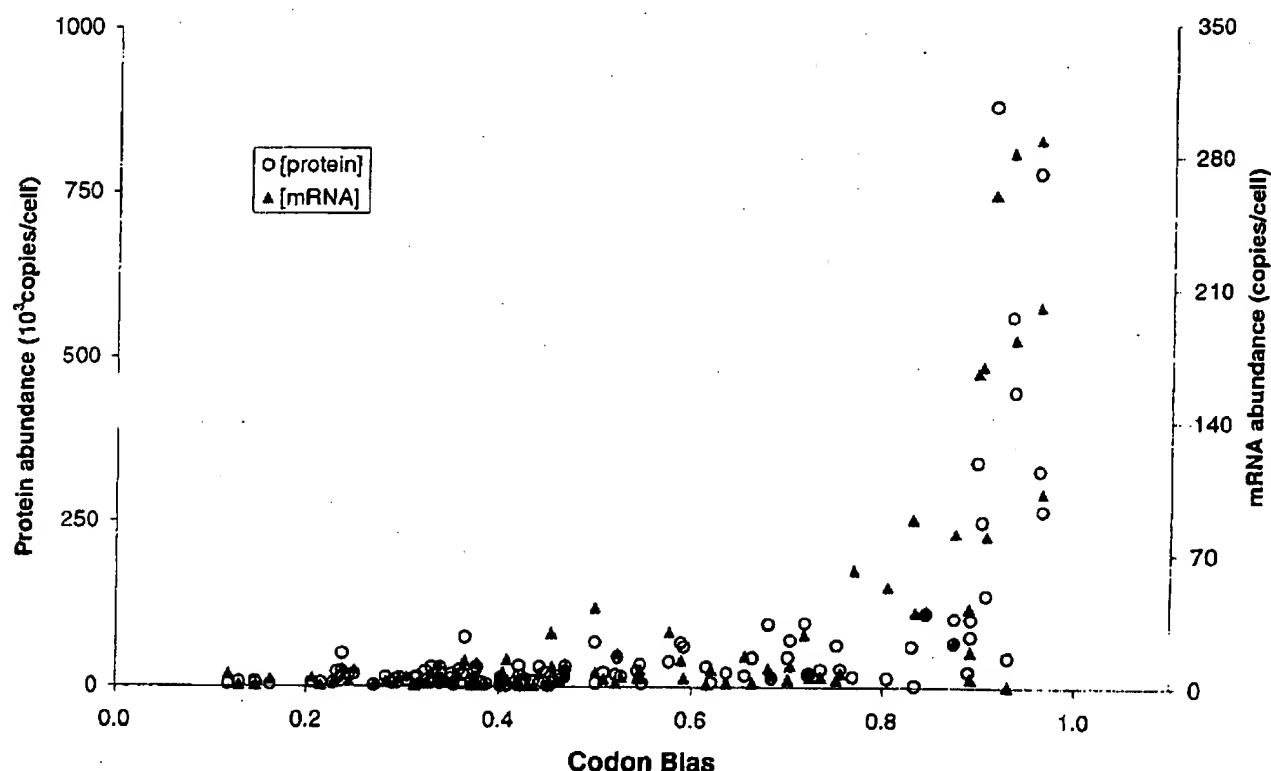


FIG. 7. Relationship between codon bias and protein and mRNA levels in this study. Yeast mRNA and protein expression levels were calculated as described in Materials and Methods. The data represent the same 106 genes as in Fig. 5.

activities calculated for each of the four proteins varied by less than 10%. Any inconsistencies in the calculation of the specific activity would result in differences in the absolute levels calculated but not in the relative numbers and would therefore not influence the correlation value determined.

The protein quantitative method used eliminates a number of potential errors inherent in previous methods for the quantitation of proteins separated by 2DE, such as preferential protein staining and bias caused by inequalities in the number of radiolabeled residues per protein. Any 2D gel-based method of quantitation is complicated by the fact that in some cases the translation products of the same mRNA migrated to different spots. One major reason is posttranslational modification or processing of the protein. Also, artifactual proteolysis during cell lysis and sample preparation can lead to multiple resolved forms of the protein. In such cases, the protein levels of spots coded for by the same mRNA were pooled. In addition, the existence of other spots coded for by the same mRNA that were not analyzed by mass spectrometry or that were below the limit of detection for silver staining cannot be ruled out. However, since this study is based on a class of highly expressed proteins, the presence of undetected minor spots below silver staining sensitivity corresponding to a protein analyzed in the study would generally cause a relatively small error in protein quantitation.

Codon bias is a measure of the propensity of an organism to selectively utilize certain codons which result in the incorporation of the same amino acid residue in a growing polypeptide chain. There are 61 possible codons that code for 20 amino acids. The larger the codon bias value, the smaller the number of codons that are used to encode the protein (19). It is

thought that codon bias is a measure of protein abundance because highly expressed proteins generally have large codon bias values (3, 13).

Nearly all of the most highly expressed proteins had codon bias values of greater than 0.8. However, we detected a number of genes with high codon bias and relative low protein abundance (Fig. 7). For example, the expressed gene with both the second largest protein and mRNA levels in the study was ENO2_YEAST (775,000 and 289.1 copies/cell, respectively). ENO1_YEAST was also present in the gel at much lower protein and mRNA levels (44,200 and 0.7 copies/cell, respectively). The codon bias values for ENO2 and ENO1 are similar (0.96 and 0.93, respectively), but the expression of the two genes is differentially regulated. Specifically, ENO1_YEAST is glucose repressed (6) and was therefore present in low abundance under the conditions used. Other genes with large codon bias values that were not of high protein abundance in the gel include EFT1, TIF1, HXK2, GSP1, EGD2, SHM2, and TAL1. We conclude that merely determining the codon bias of a gene is not sufficient to predict its protein expression level.

Interestingly, codon bias appears to be an excellent indicator of the boundaries of current 2D gel proteome analysis technology. There are thousands of genes with expressed mRNA and likely expressed protein with codon bias values less than 0.1 (Fig. 4A). In this study, we detected none of them, and only a very small percentage of the genes detected in this study had codon bias values between 0.1 and 0.2 (Fig. 4B). Indeed, in every examined yeast proteome study (5, 7, 13, 28) where the combined total number of identified proteins is 300 to 400, this same observation is true. It is expected that for the more complex cells of higher eukaryotic organisms the detection of

low-abundance proteins would be even more challenging than for yeast. This indicates that highly abundant, long-lived proteins are overwhelmingly detected in proteome studies. If proteome analysis is to provide truly meaningful information about cellular processes, it must be able to penetrate to the level of regulatory proteins, including transcription factors and protein kinases. A promising approach is the use of narrow-range focusing gels with immobilized pH gradients (IPG) (23). This would allow for the loading of significantly more protein per pH unit covered and also provide increased resolution of proteins with similar electrophoretic mobilities. A standard pH gradient in an isoelectric focusing gel covers a 7-pH-unit range (pH 3 to 10) over 18 cm. A narrow-range focusing gel might expand the range to 0.5 pH units over 18 cm or more. This could potentially increase by more than 10-fold the number of proteins that can be detected. Clearly, current proteome technology is incapable of analyzing low-abundance regulatory proteins without employing an enrichment method for relatively low-abundance proteins. In conclusion, this study examined the relationship between yeast protein and message levels and revealed that transcript levels provide little predictive value with respect to the extent of protein expression.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Science and Technology Center for Molecular Biotechnology, NIH grant T32HG00035-3, and a grant from Oxford Glycosciences.

We thank Jimmy Eng for expert computer programming, Garry Corthals and John R. Yates III for critical discussion, and Siavash Mohandesi for expert technical help.

REFERENCES

- Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. Kent. 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* 84:6970-6974.
- Aebersold, R. H., D. B. Teplow, L. E. Hood, and S. B. Kent. 1986. Electrophoresis onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. *Eur. J. Biochem.* 261:4229-4238.
- Bennetzen, J. L., and B. D. Hall. 1982. Codon selection in yeast. *J. Biol. Chem.* 257:3026-3031.
- Boucherie, H., G. Du Jardin, M. Kermorgant, C. Monribot, P. Slonimski, and M. Perrot. 1995. Two-dimensional protein map of *Saccharomyces cerevisiae*: construction of a gene-protein index. *Yeast* 11:601-613.
- Boucherie, H., F. Sagliocco, R. Joubert, I. Maillet, J. LaBarre, and M. Perrot. 1996. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* 17:1683-1699.
- Carmen, A. A., P. K. Brindle, C. S. Park, and M. J. Holland. 1995. Transcriptional regulation by an upstream repression sequence from the yeast enolase gene ENO1. *Yeast* 11:1031-1043.
- Ducet, A., I. VanOostveen, J. K. Eng, J. R. Yates, and R. Aebersold. 1998. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* 7:706-719.
- Eng, J., A. McCormack, and J. R. Yates. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5:976-989.
- Figgeys, D., A. Ducet, J. R. Yates, and R. Aebersold. 1996. Protein identification by solid phase microextraction-capillary zone electrophoresis-micro-electrospray-tandem mass spectrometry. *Nat. Biotechnol.* 14:1579-1583.
- Figgeys, D., I. VanOostveen, A. Ducet, and R. Aebersold. 1996. Protein identification by capillary zone electrophoresis/microelectrospray ionization-tandem mass spectrometry at the subfemtomole level. *Anal. Chem.* 68:1822-1828.
- Fraser, C. M., and R. D. Fleischmann. 1997. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* 18:1207-1216.
- Garrels, J. I., B. Futcher, R. Kobayashi, C. I. Latter, B. Schwender, T. Volpe, J. R. Warner, and C. S. McLaughlin. 1994. Protein identifications for a *Saccharomyces cerevisiae* protein database. *Electrophoresis* 15:1466-1486.
- Garrels, J. I., C. S. McLaughlin, J. R. Warner, B. Futcher, C. I. Latter, R. Kobayashi, B. Schwender, T. Volpe, D. S. Anderson, F. Mesquita-Fuentes, and W. E. Payne. 1997. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* 18:1347-1360.
- Gygi, S. P., and R. Aebersold. 1998. Absolute quantitation of 2-DE protein spots, p. 417-421. In A. J. Link (ed.), 2-D protocols for proteome analysis. Humana Press, Totowa, N.J.
- Harford, J. B., and D. R. Morris. 1997. Post-transcriptional gene regulation. Wiley-Liss, Inc., New York, N.Y.
- Hereford, L. M., and M. Rosbash. 1977. Number and distribution of polyadenylated RNA sequences in yeast. *Cell* 10:453-462.
- Hodges, P. E., W. E. Payne, and J. I. Garrels. 1998. The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 26:68-72.
- Klose, J., and U. Kobalz. 1995. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16:1034-1059.
- Kurland, C. G. 1991. Codon bias and gene expression. *FEBS Lett.* 285:165-169.
- Lashkari, D. A., J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* 94:13057-13062.
- Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.
- Link, A. J., L. G. Hays, E. B. Carmack, and J. R. Yates III. 1997. Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* 18:1314-1334.
- Nawrocki, A., M. K. Larsen, A. V. Podtelejnikov, O. N. Jensen, M. Mann, P. Roepstorff, A. Gorg, S. J. Fey, and P. M. Larsen. 1998. Correlation of acidic and basic carrier ampholyte and immobilized pH gradient two-dimensional gel electrophoresis patterns based on mass spectrometric protein identification. *Electrophoresis* 19:1024-1035.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007-4021.
- OWL Protein Sequence Database. 2 August 1998, posting date. [Online.] <http://bubs.gil1.leeds.ac.uk/bmb5dp/owl.html>. [8 January 1999, last date accessed.]
- Patterson, S. D., and R. Aebersold. 1995. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 16:1791-1814.
- Pennington, S. R., M. R. Wilkins, D. F. Hochstrasser, and M. J. Dunn. 1997. Proteome analysis: from protein characterization to biological function. *Trends Cell Biol.* 7:168-173.
- Shalon, D., S. J. Smith, and P. O. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6:639-645.
- Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 93:14440-14445.
- Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* 68:850-858.
- Sikorski, R. S., and P. Hieter. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122:19-27.
- Tsunasawa, S., J. W. Stewart, and F. Sherman. 1985. Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. *J. Biol. Chem.* 260:5382-5391.
- Urlinger, S., K. Kuchler, T. H. Meyer, S. Uebel, and R. Tampé. 1997. Intracellular location, complex formation, and function of the transporter associated with antigen processing in yeast. *Eur. J. Biochem.* 245:266-272.
- Varshavsky, A. 1996. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* 93:12142-12149.
- Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270:484-487.
- Velculescu, V. E., L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
- Wilkins, M. R., K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1997. Proteome research: new frontiers in functional genomics. Springer-Verlag, Berlin, Germany.
- Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann. 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379:466-469.
- Yau, J. X., M. R. Wilkins, K. Ou, A. A. Gooley, K. L. Williams, J. C. Sanchez, O. Golaz, C. Pasquali, and D. F. Hochstrasser. 1996. Large-scale amino-acid analysis for proteome studies. *J. Chromatogr. A* 736:291-302.
- YPD Website. 6 March 1998, revision date. [Online.] Proteome, Inc. <http://www.proteome.com/YPDhome.html>. [8 January 1999, last date accessed.]

Analysis of Genomic and Proteomic Data Using Advanced Literature Mining

Yanhui Hu, Lisa M. Hines, Haifeng Weng, Dongmei Zuo, Miguel Rivera,
Andrea Richardson, and Joshua LaBaer*

Institute of Proteomics, Harvard Medical School—BCMP, 240 Longwood Avenue, Boston, Massachusetts 02115

Received March 13, 2003

High-throughput technologies, such as proteomic screening and DNA micro-arrays, produce vast amounts of data requiring comprehensive analytical methods to decipher the biologically relevant results. One approach would be to manually search the biomedical literature; however, this would be an arduous task. We developed an automated literature-mining tool, termed MedGene, which comprehensively summarizes and estimates the relative strengths of all human gene–disease relationships in Medline. Using MedGene, we analyzed a novel micro-array expression dataset comparing breast cancer and normal breast tissue in the context of existing knowledge. We found no correlation between the strength of the literature association and the magnitude of the difference in expression level when considering changes as high as 5-fold; however, a significant correlation was observed ($r = 0.41$; $p = 0.05$) among genes showing an expression difference of 10-fold or more. Interestingly, this only held true for estrogen receptor (ER) positive tumors, not ER negative. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative tumors worthy of further examination.

Keywords: bioinformatics • micro-array • text mining • gene-disease association • breast cancer

Introduction

At its current pace, the accumulation of biomedical literature outpaces the ability of most researchers and clinicians to stay abreast of their own immediate fields, let alone cover a broader range of topics. For example, to follow a single disease, e.g., breast cancer, a researcher would have had to scan 130 different journals and read 27 papers per day in 1999.¹ This problem is accentuated with high-throughput technologies such as DNA micro-arrays and proteomics, which require the analysis of large datasets involving thousands of genes, many of which are unfamiliar to a particular researcher. In any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study. The ability to interpret these datasets would be enhanced if they could be compared to a comprehensive summary of what is known about all genes. Thus, there is a need to summarize existing knowledge in a format that allows for the rapid analysis of associations between genes and diseases or other specific biological concepts.

One solution to this problem is to compile structured digital resources, such as the Breast Cancer Gene Database¹ and the Tumor Gene Database.² However, as these resources are hand-curated, the labor-intensive review process becomes a rate-limiting step in the growth of the database. As a result, these

databases have a limited scale and the genes are not selected in a systematic fashion.

An alternative approach is automated text mining: a method which involves automated information extraction by searching documents for text strings and analyzing their frequency and context. This approach has been used successfully in several instances for biological applications. In most cases, it has been applied to extract information about the relationships or interactions that proteins or genes have with one another, in the literature or by functional annotation.^{3–7} Thus far, few publications have applied text-mining to examine the global relationships between genes and diseases. Perez-Iratxeta et al. automatically examined the GO (Gene Ontology) annotation of genes and their predicted chromosomal locations in order to identify genes linked to inherited disorders.⁸

To obtain a more global understanding of disease development, it would be valuable to incorporate information regarding all possible gene–disease relationships, including biochemical, physiological, pharmacological, epidemiological, as well as genetic. This information would enable comprehensive comparisons between large experimental datasets and existing knowledge in the literature. This would accomplish two things. First, it would serve to validate experiments by demonstrating that known responses occur as predicted. Second, it would rapidly highlight which genes are corroborated by the literature and which genes are novel in a given context. We have utilized a computational approach to literature mining to produce a

* To whom correspondence should be addressed: jlabae@tums.harvard.edu.

comprehensive set of gene-disease relationships. In addition, we have developed a novel approach to assess the strength of each association based on the frequency of citation and co-citation. We applied this tool to help interpret the data from a large micro-array gene expression experiment comparing normal and cancerous breast tissue.

Methods

MedGene Database. MedGene is a relational database, storing disease and gene information from NCBI, text mining results, statistical scores, and hyperlinks to the primary literature. MedGene has a web-based user-interface for users to query the database (<http://hipseq.med.harvard.edu/MedGene/>).

Text Mining Algorithms. MeSH files were downloaded from the MeSH web site at NLM (National Library of Medicine) (<http://www.nlm.nih.gov/mesh/meshhome.html>) and human disease categories were selected. LocusLink files were downloaded from the LocusLink web site at NCBI (<http://www.ncbi.nlm.gov/LocusLink/>). Official/preferred gene symbol, official/preferred gene name, and gene alternative symbols and names, all relevant annotations and URLs for each LocusLink record, were collected. Gene search terms were used for literature searching and included all qualified gene names, gene symbols, and gene family terms. Primary gene keys, predominantly qualified gene family terms and gene official/preferred symbols, were used to index Medline records. If the official/preferred gene symbols did not meet the standards to be an index, then qualified gene official/preferred names were used. A local copy of Medline records (up to July, 2002) was pre-selected.

A JAVA module examined the MeSH terms and then indexed each Medline record with the appropriate disease terms. A separate JAVA module was used to examine the titles and abstracts for gene search terms and then to index the gene-related Medline records with the relevant primary gene key(s).

Statistical Methods. For every gene and disease pair, we counted records that were indexed for both gene and disease (double positive hits), for disease only (disease single hits), for gene only (gene single hits), and for neither gene nor disease (double negative hits) to generate a 2×2 contingency table. On the basis of the contingency table-framework, we applied different statistical methods to estimate the strength of gene-disease relationships and evaluated the results. These methods included chi-square analysis, Fisher's exact probabilities, relative risk of gene, and relative risk of disease¹⁰ (<http://hipseq.med.harvard.edu/MedGene/>). In addition, we computed the "product of frequency", which is the product of the proportion of disease/gene double hits to disease single hits and the proportion of disease/gene double hits to gene single hits. To obtain a normal distribution, we transformed all the statistical scores using the natural logarithm. We selected the log of the product of frequency (LPF) to validate MedGene and to use for the analysis with the micro-array data. Spearman rank-correlation coefficients were used to assess the linear relationship between LPF and micro-array fold change in expression level.

Global Analysis. Diseases with at least 50 related genes were selected for clustering analysis, and the LPF scores were normalized with total score for each disease. Hierarchical clustering was done with the "Cluster" software and the clustering result was visualized using "TreeView" (<http://rana.lbl.gov/EisenSoftware.htm>).

Breast Tissue Micro-Arrays. Eighty-nine breast cancer samples (79% ER-positive) and 7 normal breast tissue samples were selected from the Harvard Breast SPORF frozen tissue repository and were representative of the spectrum of histological types, grades, and hormone receptor immuno-phenotypes of breast cancer. Biotinylated cRNA, generated from the total RNA extracted from the bulk tumor, was hybridized to Affymetrix U95A oligo-nucleotide micro-arrays. These micro-arrays consist of 12 400 probes, which represent approximately 9000 genes. Raw expression values were obtained using GENE-CHIP software from Affymetrix, and then further analyzed using the DNA-Chip Analyzer (dChip) custom software.

Results

Automated Indexing of Medline Records by Disease and Gene. To study the gene-disease associations in the literature, we first compiled complete lists for human diseases and human genes. To index all Medline records that were relevant to human diseases, the Medical Subject Heading (MeSH) Index of Medline records was utilized. MeSH is a controlled medical vocabulary from the National Library of Medicine and consists of a set of terms or subject headings that are arranged in both an alphabetic and an hierarchical structure. Medline records are reviewed manually and MeSH terms are added to each with software assistance.^{2,10} Twenty-three human disease category headings along with all of their child terms (see the Supporting Information, Supplemental Table 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table1.html) were selected from the 2002 MeSH Index creating a list of 4033 human diseases.

No index comparable to the MeSH Index exists for genes, and thus, it was necessary to apply a string search algorithm for gene names or symbols found in Medline text. A complete list of genes, gene names, gene symbols, and frequently used synonyms were collected from the LocusLink database at NCBI,^{11,12} which contains 53 259 independent records keyed by an official gene symbol or name (June 18th, 2002). For the purposes of this study, no distinction was made between genes and their gene products. Authors often use the same name for both, differentiating the two only by the use of italics, if at all. For the intended use of this study, this lack of distinction is unlikely to have a large effect and may in fact be beneficial.

Initial attempts to search the literature using these lists revealed several sources of false positives and false negatives (Table 1). False positives primarily arose when the searched term had other meanings, whereas false negatives arose from syntax discrepancies necessitating the development of filters to reduce these errors. The syntax issues were readily handled by including alternate syntax forms in the search terms. The false positive cases, caused by duplicative and unrelated meanings for the terms, were more difficult to manage. Where possible, case sensitive string mapping reduced inappropriate citations. In many cases, however, this was not sufficient and the terms had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.

For the purposes of data tracking, a primary gene key was selected to represent all synonyms that correspond to each gene. Medline records were indexed with a primary gene key when any synonym for that key was found in the title or abstract. Case-insensitive string mapping was used for all searches except as noted above. No additional weight was

Table 1. Systematic Sources of False Positives and False Negatives in Unfiltered Data*

source of error	error type	example	filter solution
gene symbol/name is not unique	false positive	MAC—myelin associated glycoprotein MAC—malignancy-associated protein	eliminate this term
gene symbol is unrelated abbreviation	false positive	PA—pallid homologue (mouse), pallidin (also abbrev. for Pennsylvania)	eliminate this term
gene symbol/name has language meaning	false positive	WAS—Wiskott—Aldrich Syndrome (also the word "was")	case-sensitive string search
nonstandard syntax	false negative	BAG-1 instead of BAG1	add dash term
unofficial gene name/symbol	false negative	P53 instead of TP53	add all gene nicknames
nonspecified gene name	false negative	estrogen receptor instead of Estrogen receptor 1	add family stem term

* In preliminary studies, Medline was searched for co-occurrence of genes and diseases and the resulting output was evaluated to identify error sources that were amenable to global filters. Each error source is categorized by the type of error it causes: false positives are suggested relationships that are not real and false negatives are real relationships that are underrepresented. The filter solutions used are indicated. Note that in some cases, the filter solution itself introduces error. In general, error rates maximized sensitivity, even at the expense of specificity if needed.

added for multiple occurrences of a term or the co-occurrence of multiple synonyms for the same gene key.

Medline records were searched with all qualified gene identifiers, such as the official/preferred gene symbol, the official/preferred gene name, all gene nicknames and all syntax variants. In situations where there are several members of a gene family or splice variants, some authors prefer to use a shortened gene family name, e.g., estrogen receptor, instead of estrogen receptor 1 (*ESR1*), creating a source of false negatives. For this reason, gene family stem terms were created for all genes that have an alpha or numerical suffix (e.g., *IL2RA*, *TGFB*, *ESR1*, etc.) and then used to search the literature. The family stem terms were handled separately from the specific gene names so that it would be clear when linkages were made to the gene family versus a specific member in that family.

To improve performance and accuracy, some pre-selection was applied to the records that were scanned. First, review articles were eliminated to avoid redundant treatment of citations. Second, non-English journals were removed because the natural language filters were only relevant to English publications. Finally, journals unlikely to contain primary data about gene-disease relationships were also removed (e.g., *Int. J. Health Educ.*, *Bedside Nurse*, and *J. Health Econ.*). Together, these filters reduced the 12 198 221 Medline publications (July 2002) by 37%.

Ranking the Relative Strengths of Gene-Disease Associations. In total, there were 618 708 gene-disease co-citations, in which 16% (8297) of all studied genes had been associated to a disease and 96% (3875) of all diseases had been associated to at least one gene. To rank the relative strengths of gene-disease relationships, we tested several different statistical methods and examined the results. With the exception of the relative risk estimates, the methods provided similar results with respect to the rank order of the gene-disease association strengths. However, after comparing the results to other databases and after consulting disease experts, the log of the product of frequency (LPF) was selected for further analysis because it gave the best results overall.

Validation of MedGene. In developing this tool, it was important to minimize the number of missed genes (false negatives) and misclassified genes (false positives). However, in situations when these goals were in conflict, inclusiveness was prioritized. To determine the false negative rate in MedGene, breast cancer was used as a test case because it was associated with more genes than any other human disease and because

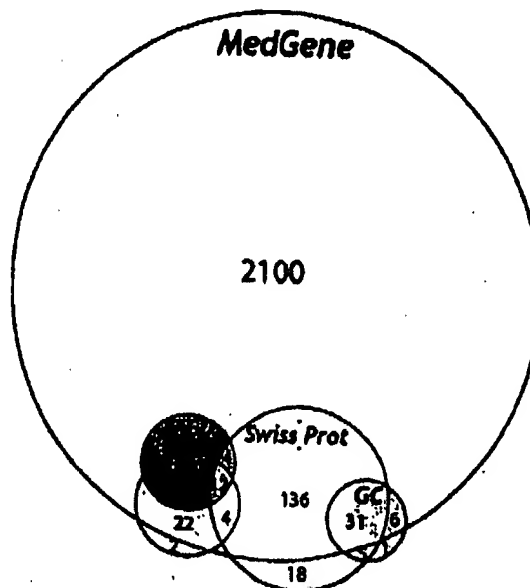


Figure 1. Estimation of the false negative rate by comparison with hand-curated databases. The breast cancer-related genes identified by MedGene were compared with those listed in several other databases including the Tumor Gene Database (TGD),² the Breast Cancer Gene Database (BCG),¹ GeneCards (GC)¹⁷ and Swissprot.¹⁸ Genes were considered false negatives if they were represented in at least one of these other databases and not in MedGene and their link to breast cancer was supported by at least one literature reference. All literature references were verified by manual review to confirm their validity. The number of genes in each database or shared by more than one database is indicated. The false negative rate was calculated by genes missed at MedGene (26)/total number of nonoverlapping genes in other databases (285).

there were several public databases that link genes to breast cancer. We compared the list of breast cancer-related genes from MedGene to these databases, illustrated in Figure 1. Among the 285 distinct breast cancer-related genes that were supported by at least one literature citation in these hand-curated databases, 26 were absent from MedGene, suggesting a false negative rate of approximately 9%. To determine why these were missed, all literature references for these genes (80

papers) were reviewed manually (see the Supporting Information, Supplemental Table 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_2.html). Among these papers, most false negatives were caused by nonstandard gene terms or gene terms eliminated by our specificity filters. Few genes were missed because they were only mentioned in review papers (0.4%) or they appeared only in the body of the manuscript but not the abstract or title (1.1%). Of note, MedGene identified approximately 2000 additional breast cancer-related genes not listed in any other database.

To assess the false positive error rate, two complementary approaches were used: a detailed analysis of one disease and a global examination of 1000 diseases. The detailed approach examined the false positive error rate and its sources, whereas the global approach tested whether the overall results made biomedical sense.

Using the LPP, 1467 genes related to prostate cancer were assembled in rank order. We then retrieved approximately 300 Medline records each for the highest ranked 100 and the lowest ranked 200 genes and manually reviewed the titles and abstracts to determine the verity of the association. Nearly 80% of the highest ranked 100 genes fell into one of the five categories that reflect meaningful gene-disease relationships (see the Supporting Information, Supplemental Table 3, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_3.html). Among the lowest ranked 200 genes, approximately 70% reflected true relationships. Of the 600 records reviewed, there were only two in which the association between the gene and the disease was described as negative. Both were genes with very low scores. In both cases, the authors did not argue the absence of any relationship, but rather that a particular feature of the gene or protein was not shown to be related to human prostate cancer.^{13,14}

The coincidence of some gene symbols with medical abbreviations, chemical abbreviations and biological abbreviations resulted in most of the false positives (see the Supporting Information, Supplemental Table 4, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table_4.html), emphasizing the importance of the filters that were added in the search algorithm (Table 1). Without the filters, the false positive rate more than doubled, and the false negative rate rose dramatically (data not shown). For example, among the papers about breast cancer, there were only 12 Medline records that referred to *ESR1* and 10 to *ESR2*, whereas almost 2000 papers mentioned estrogen receptor without specifying *ESR1* or *ESR2*; this latter group was detected by the family term filter.

To further validate these results, a global analysis of the gene-disease relationships described by MedGene was performed. For this experiment, it was reasoned that the more closely related the diseases are to one another, the more they will be related to the same gene sets. Thus, if the relationships defined by MedGene accurately reflected the literature, then an unsupervised hierarchical clustering of the gene data should group diseases in a manner consistent with common medical thinking. Conversely, if the clustered diseases do not make sense biologically or medically, it may reflect excessive false positives, false negatives, or inappropriate scoring of the data.

To execute this experiment, the gene sets and the corresponding LPP values for 1000 randomly selected diseases (each with at least 50 gene relationships) were used as a dataset for clustering the diseases. A review of the results showed that the resulting disease clusters were indeed logical based upon common medical knowledge (see the Supporting Information,

Supplemental Figure 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html). For example, in one such cluster shown in Figure 2, diabetes and its complications grouped together and were also closely linked to diseases associated with starvation states.

The number of genes associated with a given disease can be estimated by adjusting the MedGene number up by the false negative rate (~9%) and down by the false positive rate (~26% on average). Using this, the average disease has 103.7 ± 45.3 (mean \pm s.d.) genes associated with it, although the range is quite broad with 2359 genes related to breast cancer, 2122 genes related to lung cancer and no genes related to a number of diseases.

Applying MedGene to the Analysis of Large Datasets. Access to a comprehensive summary of the genes linked to human diseases provided an opportunity to analyze data obtained from a high-throughput experiment. We compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples. Micro-array analysis identified 2286 genes that had greater than a 1-fold difference in mean expression level between breast cancer samples and normal breast samples. Using MedGene, we sorted the 2286 genes into four classes: 555 genes directly linked to breast cancer in the literature by gene term search (first-degree association by gene name); 328 genes directly linked by family term search (first-degree association by family term); 1021 genes linked to breast cancer only through other breast cancer genes (second-degree association); and 505 genes not previously associated with breast cancer. (See the Supporting Information, Supplemental Figure 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_2.html.) Among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease. Among the remaining 38 genes, 9 had been related to other cancers, specifically esophageal, colon, uterine, skin, and cervix.

To determine whether the genes highlighted by the micro-array analysis were more likely to have been previously linked to breast cancer in the literature, we created a two-dimensional plot of the fold change of expression level between breast cancer and normal tissue versus the literature score (LPP) (Figure 3A). There was a broad spread of expression changes among the genes directly linked to breast cancer ranging from less than 1-fold change (68%) to over 40-fold (0.3%). Notably, the majority of genes with greater than 10-fold expression changes were linked to breast cancer by first-degree association.

Among all 754 genes directly linked to breast cancer in the literature, there was no correlation between LPP and micro-array fold change ($r = 0.018$, p -value = 0.62). However, when we stratified the analysis based on the magnitude of the fold change, we observed an increasing trend in correlation (Figure 3B) suggesting that genes with a more substantial change in expression level were more likely to have a stronger association in the literature. For genes that had 10-fold change or more in expression level, the correlation increased to 0.41 (p -value = 0.05).

When we evaluated the micro-array data separately for ER positive and ER negative tumors, the trend in correlation between fold change and literature score was highly dependent on estrogen receptor status. Interestingly, there was a similar trend in correlation for ER positive tumors, but no trend in correlation for ER negative tumors.

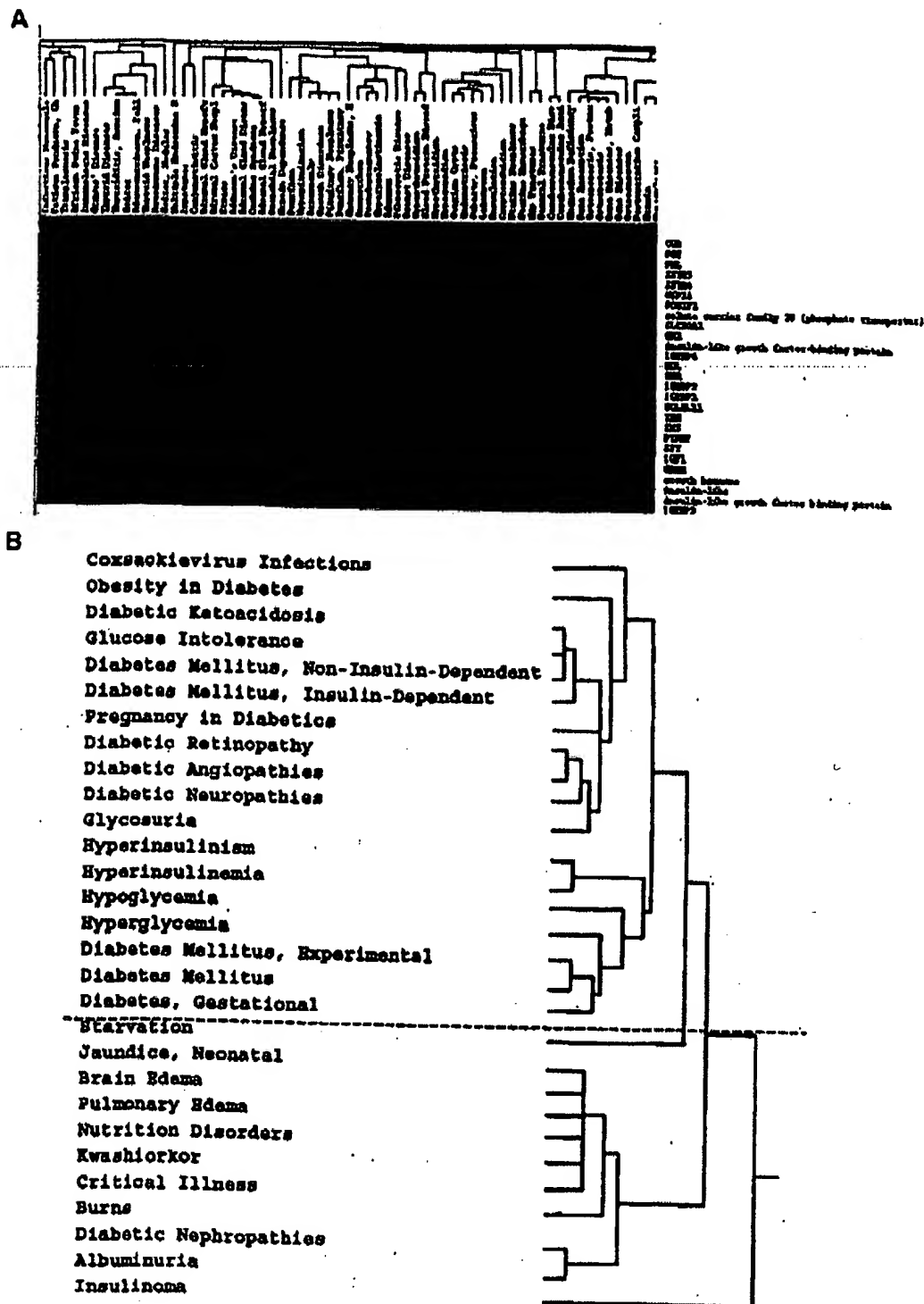


Figure 2. Global validation by clustering analysis. 2(A). The gene sets and the corresponding LPF values for 1000 diseases, each with at least 50 gene relationships, were used in an unsupervised clustering of the diseases based on the gene patterns associated with them. A sample of the data is shown here. 2(B). One of the resulting clusters is shown that corresponds to blood sugar states. Diabetes terms (above the line) and starvation states terms (under the line) clustered together. Within these groups, there is also clustering of diabetic small vessel complications, altered serum chemistries, nutritional disorders, etc. (Supplemental Figure 1: <http://hipseq.med.harvard.edu/MedGene/publications/Figure 1.html>).

Finally, to validate our findings, we computed similar correlations between the breast cancer expression data and LPF scores generated by MedGene for hypertension, a

disease unrelated to breast cancer. As expected, we did not observe an increasing trend in correlation for hypertension.

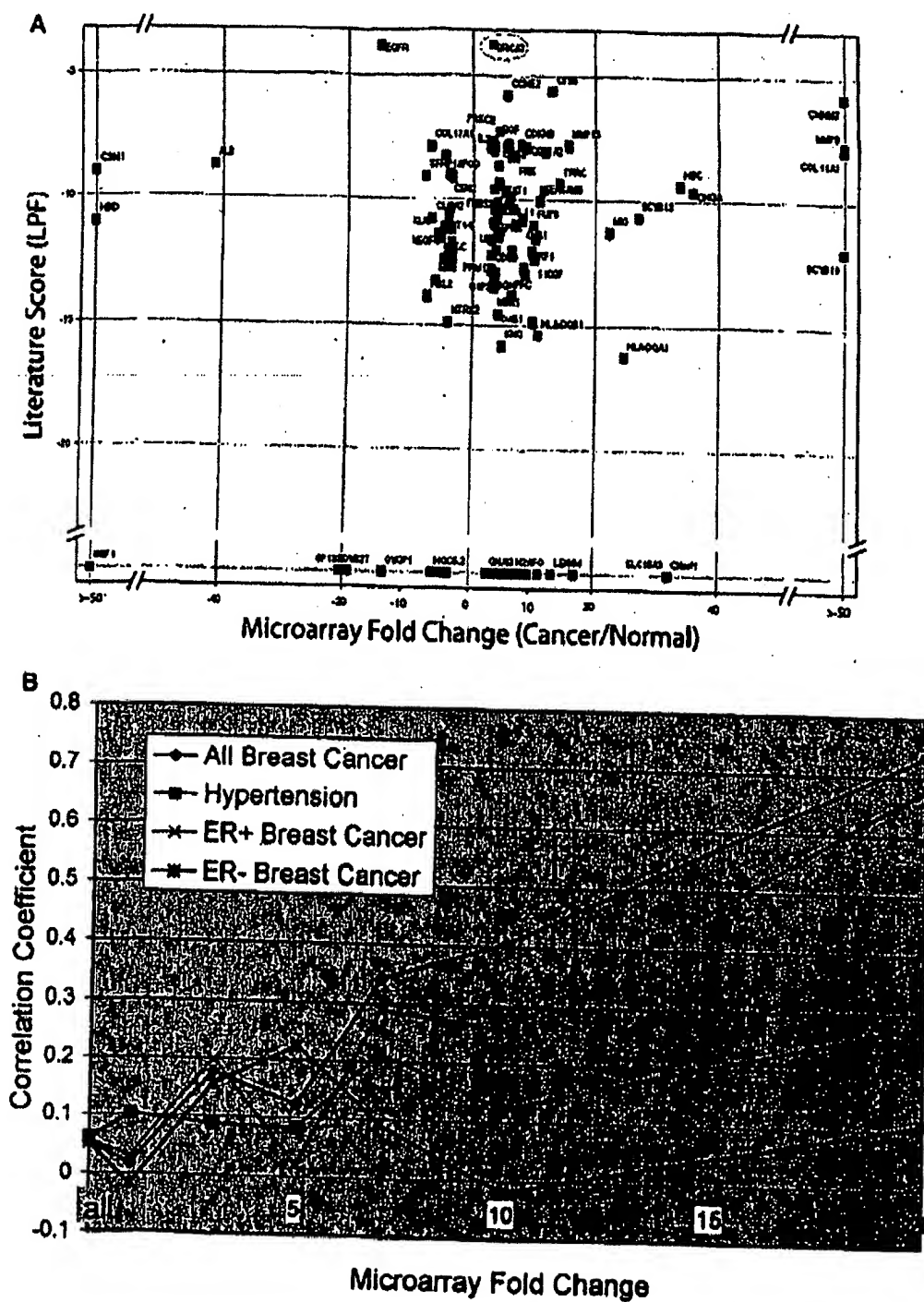


Figure 3. Relationship between literature score and functional data for breast cancer. **3A.** The data from an expression analysis of samples for breast tumors and normal breast tissue were analyzed to indicate the fold difference of expression level between breast tumor and normal sample (cutoff ≥ 3 -fold change). The fold changes were plotted against the literature score for the same gene set. Green dots represent first-degree association by gene search, blue dots represent first-degree association by family search and red dots represent no-association. Some well-studied genes, such as BRCA2 (pink circle), are not reflected by a substantial difference in expression level. Furthermore, the majority of genes that have no association with breast cancer in the literature had less than 10-fold of expression level between tumor and normal breast samples (y-axis) in relation to the amount of fold change of expression level (x-axis). Gene rank lists were generated for breast cancer (blue) and hypertension (pink). Correlations were also computed between the breast cancer gene LPF scores and fold change expression data among estrogen receptor positive tumors only (light blue) and estrogen receptor negative tumors only (purple).

Table 2. Top 25 Genes Related to Selected Human Diseases*

breast neoplasms	hypertension	rheumatoid arthritis	bipolar disorder	atherosclerosis
estrogen receptor	<i>REN</i>	<i>RA</i>	<i>ERDA1</i>	apolipoprotein
<i>PCR</i>	<i>DBP</i>	<i>TNFRSF10A</i>	<i>SNAP29</i>	<i>APOE</i>
<i>ERBB2</i>	<i>LEP</i>	<i>CRP</i>	<i>PFKL</i>	<i>LDLR</i>
<i>BRCA1</i>	<i>AGT</i>	<i>AS</i>	<i>DRD2</i>	<i>ELN</i>
<i>BRCA2</i>	<i>INS</i>	<i>ESR1</i>	<i>TRH</i>	<i>ARC1</i>
<i>EGFR</i>	kalikrein	<i>HLA-DRB1</i>	<i>IMPA2</i>	<i>APOB</i>
<i>CYP19</i>	<i>ACE</i>	<i>DR1</i>	<i>HTR3A</i>	<i>APOA1</i>
<i>TFF1</i>	endothelin	interleukin	<i>DRD3</i>	<i>MSR1</i>
<i>PSEN2</i>	<i>S100A6</i>	<i>TNF</i>	<i>REM</i>	<i>LPL</i>
<i>TP53</i>	<i>BDK</i>	<i>IL6</i>	<i>KCNN3</i>	<i>PON1</i>
<i>CES3</i>	<i>DIAPH</i>	collagen	<i>DRD4</i>	plasminogen
<i>CEACAM5</i>	<i>SAR1</i>	<i>IL1A</i>	<i>HTR2C</i>	activator inhibitor
<i>ERBB3</i>	<i>PIH</i>	<i>ACR</i>	<i>RELN</i>	<i>PLG</i>
cyclin	<i>CD59</i>	<i>TNFRSF12</i>	<i>DBH</i>	vascular cell
<i>COX5A</i>	<i>ALB</i>	<i>IL2</i>	<i>MAOA</i>	adhesion molecule
cathepsin	<i>CYP11B2</i>	<i>CHI3L1</i>	<i>COMT</i>	<i>ATOH1</i>
<i>ERBB4</i>	<i>MAT2B</i>	<i>IL8</i>	<i>HTR2A</i>	<i>VWF</i>
<i>TRAM</i>	angiotensin receptor	interleukin 1 matrix	<i>SYNJ1</i>	<i>INS</i>
<i>CCND1</i>	<i>AGTR2</i>	metalloproteinase	<i>INPP1</i>	<i>ARG2</i>
<i>ECF</i>	<i>NPPA</i>	interferon	<i>NEDD4L</i>	<i>ABCA1</i>
<i>MUC1</i>	<i>LVM</i>	<i>CD68</i>	<i>FRA13C</i>	<i>OLR1</i>
insulin-like	<i>DBH</i>	<i>IL4</i>	transducer of	collagen
<i>BCL2</i>	<i>NPY</i>	<i>IL17</i>	<i>ERBB2</i>	<i>MCP</i>
mucin	<i>POMC</i>	<i>MMP3</i>	<i>BATF3</i>	lipoprotein
<i>FGF3</i>	neuropeptide	<i>SIL</i>	<i>ATP1B3</i>	<i>APOA2</i>
			<i>DRD5</i>	intercellular
				adhesion molecule
				<i>RAB27A</i>

* MedGene results for the top 25 genes associated with breast neoplasms, hypertension, rheumatoid arthritis, bipolar disorder, and atherosclerosis, respectively, ranked by LPP scores. The hyperlink to all the papers co-citing the gene and the disease is available at MedGene website (<http://hipseq.med.harvard.edu/MedGene/>).

Discussion

The Human Genome Project heralded a new era in biological research where the emphasis on understanding specific pathways has expanded to global studies of genomic organization and biological systems. High-throughput technologies can provide novel insight into comprehensive biological function but also introduces new challenges. The utility of these technologies is limited to the ability to generate, analyze, and interpret large gene lists. MedGene, a relational database derived by mining the information in Medline, was created to address this need. MedGene users can query for a rank-ordered list of human gene-disease relationships (Table 2) for one or more diseases. Each entry is hyperlinked to the original papers supporting each association and to other relevant databases.

MedGene is an innovative extension of previous text mining approaches. Perez-Iratxeta et al. used the GO annotation and their chromosomal locations to predict genes that may contribute to inherited disorders.⁸ MedGene takes a broader view and includes all diseases and all possible gene-disease relationships. Furthermore, MedGene utilizes co-citation to indicate a relationship rather than GO annotation, which is limited to the subset of genes that have GO annotation. Our approach is complementary to that taken by Chaussabel and Sher, who used the frequency of co-cited terms to cluster genes into a hierarchy of gene-gene relationships.⁹

A unique aspect of this tool is the ability to assess the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation. This presupposes that most co-citations describe a positive association, often referred to as publication bias¹⁰ and is supported by our observations

that negative associations are rare (Supplemental Table 3; <http://hipseq.med.harvard.edu/MedGene/publication/s-Table3.html>). Of course, relationships established by frequency of co-citation do not necessarily represent a true biological link; however, it is strong evidence to support a true relationship.

Another important feature of MedGene is the implementation of software filters that substantially reduced the error rate. We estimate that less than 10% of all associations were missed and at least 70% of even the weakest associations were real. For this study, all of the filters that we applied were general ones, e.g., expanding the list of all gene names to address the different syntax forms used by different journals, eliminating gene names that correspond to common English words, etc. The majority of the remaining search term ambiguities were idiosyncratic and difficult to identify systematically without causing a significant rise in false negatives. Alternative approaches, such as the examination of the nearest neighbor terms, need to be considered to further reduce the false positive rate.

It is not uncommon to see expression changes in microarray experiments as small as 2-fold reported in the literature. Even when these expression changes are statistically significant, it is not always clear if they are biologically meaningful. When comparing expression levels of disease to normal tissue, one expects an enrichment of known disease-related genes to appear in the altered expression group. MedGene provided a unique opportunity to test this notion in the context of existing knowledge on a novel breast cancer microarray dataset. For genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a known role in the disease. This

Table 3. Genes with Large Expression Changes in ER- but Not in ER+ Breast Tumors

gene symbol	fold change (ER+)	fold change (ER-)
KRTHB1	1.0	610.8
BRSS3	1.2	89.4
DKK1	1.2	69.8
ZIC1	1.9	59.6
TLR1	1.0	38.5
KIAA0680	2.6	33.2
CDKN3	1.0	30.6
EBI2	4.0	27.9
GZMB	3.8	21.9
STK18	4.7	18.6
GPR49	1.0	14.6
MYO10	1.6	14.4
LAD1	-1.0	13.5
POLE2	4.2	13.0
HMG4	4.4	12.9
BCL2L11	-1.2	12.3
LRP8	2.9	12.2
CCNB2	1.0	11.8
CCNE2	4.0	11.6
FGB	-4.3	11.1
KNSL6	2.9	10.9
HIF5	3.0	10.2
SERPINH2	4.6	10.2
YAP1	1.0	10.0
LPFB	-1.3	-10.4
TCEA2	-1.1	-10.8
TFF1	1.3	-11.4
COL17A1	-4.1	-15.7
POPS	1.1	-16.2
BPAG1	-4.6	-22.3
PDZK1	-1.1	-36.8
VECFB	-2.8	-51.5
MUC6	-1.4	-64.9
SERPINA5	-1.0	-83.1
MEIS1	-1.6	-85.9
CA12	2.4	-150.3

Table 3. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative, but not ER positive breast tumors. All of these genes have either never been co-cited with breast cancer or have a weak association except those marked with an *.

reflects the many genes whose role in breast cancer may not involve large changes in expression in sporadic tumors (e.g., *BRCA1* and *BRCA2*) and genes whose modest changes in expression may be unrelated to the disease. Strikingly, among genes with a 10-fold change or more in expression level, there was a strong and significant correlation between expression level and a published role in the disease, providing the first global validation of the micro-array approach to identifying disease-specific genes.

The results derived from MedGene have two implications. First, a careful hunt for corroborating evidence of a role in breast cancer should precede any further study of genes with less than 5-fold expression level changes. Second, any genes with 10-fold changes or more are likely to be related to breast cancer and warrant attention. It is likely that this threshold will change depending on the disease as well as the experiment.

Interestingly, the observed correlation was only found among ER-positive tumors, not ER-negative. This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently. The MedGene approach identified a set of relatively understudied, yet highly expressed genes in ER-negative tumors that are worthy of further examination (Table 3).

In conclusion, we have developed an automated method of summarizing and organizing the vast biomedical literature. To our knowledge, the resulting database is the most comprehensive and accurate of its kind. By generating a score that reflects the strength of the association, it provides an important tool for the rapid and flexible analysis of large datasets from various high-throughput screening experiments. Furthermore, it can be used for selecting subsets of genes for functional studies, for building disease-specific arrays, for looking at genes common to multiple diseases and various other high-throughput applications. In the future, it will be possible to enhance the utility of the MedGene database by building links between genes and other MeSH terms as well as other biological processes and concepts, such as cell division and responses to small molecules.

Acknowledgment. We would like to thank P. Braun, L. Garraway, J. Pearlberg, and other members of our institute for helpful discussion. Many thanks to the NLM (National Library of Medicine) for licensing of MEDLINE and the annotation effort of adding MeSH indexes for MEDLINE abstracts. This work was funded by grants from the Breast Cancer Research Foundation and an NHLBI PGA Grant (Vol HL66582-02).

Supporting Information Available: Twenty-three human disease category headings along with all of their child terms selected from the 2002 MeSH Index (Supplemental Table 1); analysis of the causes of false negatives in MedGene (Supplemental Table); meaningful gene-disease relationships found in MedGene (Supplemental Table 3); causes for incorrect assignment of gene indexes (Supplemental Table 4); a review of the results, showing that the resulting disease clusters were indeed logical (Supplemental Figure 1); and a review of the results showing that among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease (Supplemental Figure 2). This material is available free of charge via the Internet at <http://pubs.acs.org> and at the web sites mentioned in the text.

References

- (1) Baasiri, R. A.; Glasser, S. R.; Steffen, D. L.; Wheeler, D. A. *Oncogene* 1999, 18, 7958-7965.
- (2) Steffen, D. L.; Levine, A. E.; Yarus, S. L.; Baasiri, R. A.; Wheeler, D. A. *Bioinformatics* 2000, 16, 639-649.
- (3) Marcotta, E. M.; Xenarios, I.; Eisenberg, D. *Bioinformatics* 2001, 17, 359-363.
- (4) Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. *Bioinformatics* 2001, 17, 155-161.
- (5) Jernsen, T. K.; Laegreid, A.; Komorowski, J.; Hovig, E. *Nat. Genet.* 2001, 28, 21-28.
- (6) Chaussabel, D.; Sher, A. *Genome Biol.* 2002, 3, RESEARCH0055.
- (7) Gibbons, F. D.; Roth, P. P. *Genome Res.* 2002, 12, 1574-1581.
- (8) Perez-Iratxe, C.; Bork, P.; Andrade, M. A. *Nat. Genet.* 2002, 31, 316-319.
- (9) Funk, M. E.; Reid, C. A. *Bull. Med. Libr. Assoc.* 1983, 71, 176-183.
- (10) Humphrey, S. M.; Miller, N. E. *J. Am. Soc. Inf. Sci.* 1987, 38, 184-196.
- (11) Maglott, D. R.; Katz, K. S.; Scotte, H.; Pruitt, K. D. *Nucleic Acids Res.* 2000, 28, 126-128.
- (12) Pruitt, K. D.; Maglott, D. R. *Nucleic Acids Res.* 2001, 29, 137-140.
- (13) Wadelius, M.; Andersson, A. O.; Johansson, J. E.; Wadelius, C.; Rane, E. *Pharmacogenetics* 1999, 9, 333-340.
- (14) Adam, R. M.; Borer, J. G.; Williams, J.; Eastham, J. A.; Loughlin, K. R.; Freeman, M. R. *Endocrinology* 1999, 140, 5866-5875.
- (15) Montori, V. M.; Smieja, M.; Guyatt, G. H. *Mayo Clin. Proc.* 2000, 75, 1284-1288.
- (16) Denenberg, V. H. *Statistics Experimental Design for Behavioral and Biological Research*; Wiley-Liss: New York, 1978.
- (17) Rebhan, M.; Challa-Caspi, V.; Priusky, J.; Lancet, D. *Trends Genet.* 1997, 13, 183.
- (18) Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* 2000, 28, 45-48. PR0340227

SECOND DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I am currently employed by Genentech, Inc. where my job title is Staff Scientist.
2. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As I stated in my previous Declaration dated May 7, 2004 (attached as Exhibit A), my laboratory has been employing a variety of techniques, including microarray analysis, to identify genes which are differentially expressed in human tumor tissue relative to normal human tissue. The primary purpose of this research is to identify proteins that are abundantly expressed on certain human tumor tissue(s) and that are either (i) not expressed, or (ii) expressed at detectably lower levels, on normal tissue(s).
4. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor tissue at significantly higher levels than in normal human tissue. To date, we have successfully generated antibodies that bind to 31 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human tumor tissue and normal tissue. We have then quantitatively compared the levels of mRNA and protein in both the tumor and normal tissues analyzed. The results of these analyses are attached herewith as Exhibit B. In Exhibit B, "+" means that the mRNA or protein was detectably overexpressed in the tumor tissue relative to normal tissue and "-" means that no detectable overexpression was observed in the tumor tissue relative to normal tissue.
5. As shown in Exhibit B, of the 31 genes identified as being detectably overexpressed in human tumor tissue as compared to normal human tissue at the mRNA level, 28 of them (i.e., greater than 90%) are also detectably overexpressed in human tumor tissue as compared to normal human tissue at the protein level. As such, in the cases where we have been able to quantitatively measure both (i) mRNA and (ii) protein levels in both (i) tumor tissue and (ii) normal tissue, we have observed that in the vast majority of cases, there is a very strong correlation between increases in mRNA expression and increases in the level of protein encoded by that mRNA.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4-5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor tissue relative to a normal tissue more often than not correlates to a similar increase in abundance of the encoded protein in the tumor tissue relative to the normal tissue. In fact, it remains a generally accepted working assumption in molecular biology that increased mRNA levels are more often than not predictive of elevated levels of the encoded protein. In fact, an entire industry focusing on the research and development of therapeutic antibodies to treat a variety of human diseases, such as cancer, operates on this working assumption.
7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 3-29-00

By: Paul Polakis

Paul Polakis, Ph.D.

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South San Francisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of
Biochemistry, Michigan State University
East Lansing, Michigan

PUBLICATIONS:

1. Polakis, P. G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.
2. Polakis, P.G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.
3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.
4. Uhing, R.J., Polakis, P.G. and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.
5. Polakis, P.G., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.
6. Uhing, R. J., Dillon, S., Polakis, P. G., Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.
7. Polakis, P.G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.
8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.
9. Polakis, P., Weber, R.F., Nevins, B., Didsbury, J. Evans, T. and Snyderman, R. 1989 Identification of the *ral* and *rac1* Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.
10. Snyderman, R., Perianin, A., Evans, T., Polakis, P. and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., Polakis, P.G., Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. *J. Biol. Chem.* 265, 5990-6001.
12. Yatani, A., Okabe, K., Polakis, P., Halenbeck, R., McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K⁺ Channels. *Cell*. 61, 769-776.
13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and Polakis, P.G. 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. *Mol. Cell. Biol.* 10, 5977-5982.
14. Polakis, P.G., Rubinfeld, B., Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. *Proc. Natl. Acad. Sci. USA* 88, 239-243.
15. Moran, M. F., Polakis, P., McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. *Mol. Cell. Biol.* 11, 1804-1812.
16. Rubinfeld, B., Wong, G., Bekesi, E., Wood, A., McCormick, F. and Polakis, P. G. 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21^{ras} Stimulation and Promotes Guanine Nucleotide Exchange. *Internatl. J. Peptide and Prot. Res.* 38, 47-53.
17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and Polakis, P. 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21^{rap1}. *Cell* 65, 1033-1042.
18. Zhang, K., Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., Polakis, P., McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. *Science* 254, 1630-1634.
19. Martin, G., Yatani, A., Clark, R., Polakis, P., Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21^{ras}-dependent Inhibition of Muscarinic Atrial K⁺ Channel Currents. *Science* 255, 192-194.
20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and Polakis, P. 1992 Regulation of p21ras by GTPase Activating Proteins. *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 56, 237-241.
21. Pronk, G. B., Polakis, P., Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60^{v-src} Can Associate with and Phosphorylate the p21^{ras} GTPase Activating Protein. *Oncogene* 7, 389-394.
22. Polakis P. and McCormick, F. 1992 Interactions Between p21^{ras} Proteins and Their GTPase Activating Proteins. In Cancer Surveys (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* 69, 551-558.
24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34^{cdc2} Kinase in vitro. *J. Biol. Chem.* 267, 10780-10785.
25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21^{rap1A} and Components of the Budding pathway of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4084-4092.
26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. *Mol. Cell. Biol.* 12, 4634-4642.
27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nuno, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. *J. Biol. Chem.* 267, 25709-25713.
28. Janoueix-Lerosey, I., Polakis, P., Tavittian, A. and deGuznberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. *Biochem. Biophys. Res. Commun.* 189, 455-464.
29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.
30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21^{ras} with GAP, exchange factors, and its biological effector target. *J. Biol. Chem.* 268, 9157-9160.
31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta-catenin. *Science* 262, 1731-1734.
32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F., Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. *Cytogenet. Cell Genet.* 66, 18-21.
33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. *Cancer Res.* 54, 552-559.
34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavittian, A., Polakis, P. and DeGuznberg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. *Biochem. Biophys. Res. Commun.* 202, 967-975.
35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. *Cancer Res.* 54, 3676-3681.

36. Rubinfeld, B. and Polakis, P. 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis P. 1995 The APC protein and E-cadherin form similar but independent complexes with α -catenin, β -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. 1995 Regulation of intracellular β -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., Polakis, P. McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and Polakis, P. 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and Polakis, P. 1996 Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and Polakis, P. 1996 Deletion of amino-terminal structure stabilizes β -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and Polakis P. 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
-
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., Polakis, P. and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and Polakis P. 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., Polakis, P., Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. *Proc. Natl. Acad. Sci.* 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., Polakis, P. and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. *J. Cell. Biol.* 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and Polakis, P. 1997 Stabilization of β -catenin by genetic defects in melanoma cell lines. *Science* 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 *Biochem. Biophys. Acta*, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and Polakis, P. 1997 Loss of β -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. *Cancer Res.* 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and Polakis, P. 1997 Induction of a β -catenin-LEF-1 complex by wnt-1 and transforming mutants of β -catenin. *Oncogene* 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and Polakis P., 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. *Cell Growth Differ.* 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and Polakis P., 1998 Down regulation of β -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β -catenin and glycogen synthase kinase 3 β . *Current Biology* 8, 573-581.

56. Polakis, P. 1998 The oncogenic activation of β -catenin. *Current Opinions in Genetics and Development* 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and Paul Polakis. 1999 The F-box protein β -TrCP associates with phosphorylated β -catenin and regulates its activity in the cell. *Current Biology* 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, Paul Polakis and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β -catenin transactivation in intestinal tumours. *Oncogene* 18, 2883-91.

59. Meng J, Glick JL, Polakis P, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. *J Biol Chem.* 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, Paul Polakis and Steve Byers 1999 The ubiquitin-proteasome pathway and serine kinase activity modulate APC mediated regulation of β -catenin-LEF signaling. *J. Biol. Chem.* 274(23):16641-5.
61. Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. *Adv Exp Med Biol.* 470, 23-32
62. Shen Z, Batzer A, Koehler JA, Polakis P, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. *Oncogene.* 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, Polakis P, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. *FEBS Lett.* 458, 247-51.
65. Peifer M, Polakis P. 2000 Wnt signaling in oncogenesis and embryogenesis—a look outside the nucleus. *Science* 287,1606-9.
66. Polakis P. 2000 Wnt signaling and cancer. *Genes Dev*;14, 1837-1851.
67. Spink KE, Polakis P, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. *EMBO J* 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., Polakis, P. and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. *Cancer Res* 61, 4197-4204.
69. Rubinfeld B, Tice DA, Polakis P. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. *J Biol Chem* 276, 39037-39045.
-
70. Polakis P. 2001 More than one way to skin a catenin. *Cell* 2001 105, 563-566.
71. Tice DA, Soloviev I, Polakis P. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. *J Biol. Chem.* 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. *J Biol Chem.* 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. *Curr. Biol.* 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. *Cancer Res.* 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. *Exp. Cell Res.* 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIs—quantitative gene expression profiling in normal and cancer tissues. *Bioinformatics*, April 8.

EXHIBIT B

	tumor mRNA	tumor IHC
UNQ2525	+	+
UNQ2378	+	+
UNQ972	+	-
UNQ97671	+	+
UNQ2964	+	+
UNQ323	+	+
UNQ1655	+	+
UNQ2333	+	+
UNQ9638	+	+
UNQ8209	+	+
UNQ6507	+	+
UNQ8196	+	+
UNQ9109	+	+
UNQ100	+	+
UNQ178	+	+
UNQ1477	+	+
UNQ1839	+	+
UNQ2079	+	+
UNQ8782	+	+
UNQ9646	+	-
UNQ111	+	+
UNQ3079	+	+
UNQ8175	+	+
UNQ9509	+	+
UNQ10978	+	-
UNQ2103	+	+
UNQ1563	+	+
UNQ16188	+	+
UNQ13589	+	+
UNQ1078	+	+
UNQ879	+	+

MOLECULAR BIOLOGY OF **THE CELL** THIRD EDITION

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson



Garland Publishing, Inc.
New York & London

GARLAND STAFF

Text Editor: Miranda Robertson

Managing Editor: Ruth Adams

Illustrator: Nigel Orme

Molecular Model Drawings: Kate Hesketh-Moore

Director of Electronic Publishing: John M. Roblin

Computer Specialist: Chuck Bartelt

Disk Preparation: Carol Winter

Copy Editor: Shirley M. Cobert

Production Editor: Douglas Goertzen

Production Coordinator: Perry Bessas

Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge.

Julian Lewis received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College, London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts . . . [et al.] .—3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0 8153 1619 1 (hard cover) ISBN 0 8153 1620 8 (pbk.)

1. Cytology. 2. Molecular biology. I. Alberts, Bruce.

II. [et al.] 1. Cells. 2. Molecular biology. Q11531.2 M713 1994

QH531.2.M64 1994

574.37—dc20

DNLM/DLC

for Library of Congress

93-05907

CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover: The photograph shows a rat nerve cell in culture. It is labeled with a fluorescent antibody that stains its cell body and dendritic processes (*red*). Nerve terminals (*green*) from other neurons not visible, which have made synapses on the cell, are labeled with a different antibody (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gayn Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoc author Bruce Alberts and famous mountaineer guide Mugs Stump (1910-1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Oliver.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals³

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein⁴

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9-2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

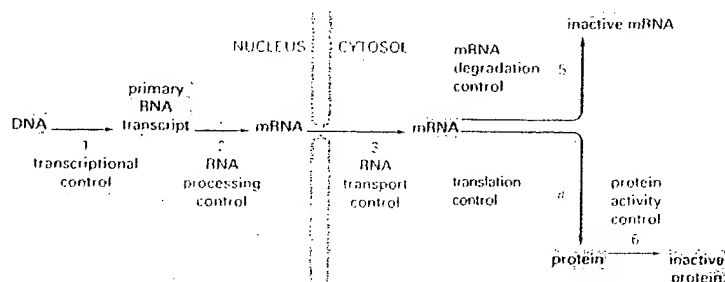


Figure 9-2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

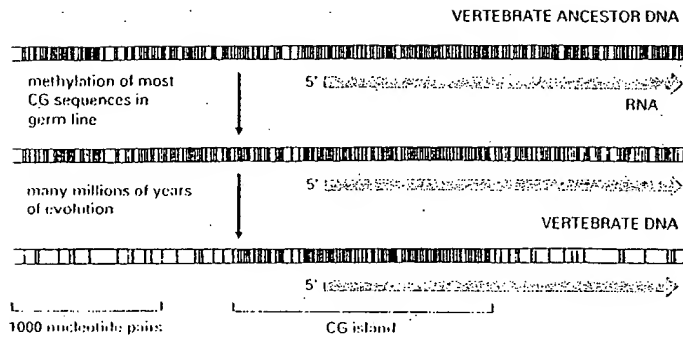


Figure 9-71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A black line marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a red line marks the location of a methylated CG dinucleotide.

Summary

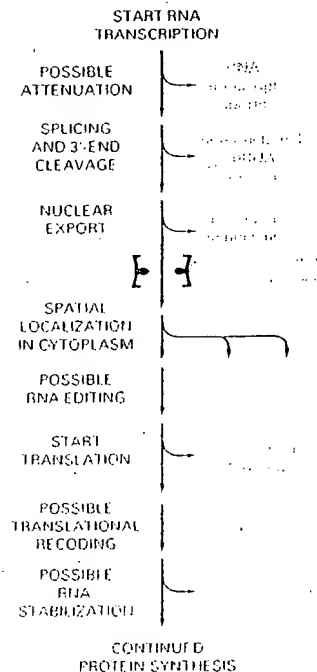
The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Prokaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these posttranscriptional controls, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than transcriptional control, for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9-72).



Possible post-transcriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

 Garland Science
Taylor & Francis Group

Garland

Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilemia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Cell Biology Interactive

Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

© 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.
© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts ... [et al.]. -- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
IDNLM: 1. Cells. 2. Molecular Biology. I
QH1581.2.M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group,
22 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860-921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Evan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila*, *Arabidopsis*. Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s. © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert I. Lehninger, unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

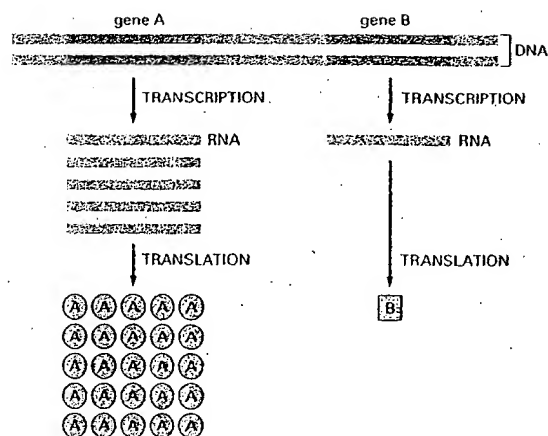


Figure 6-3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6-3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription**.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6-4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribonucleic acid*) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6-5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6-6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

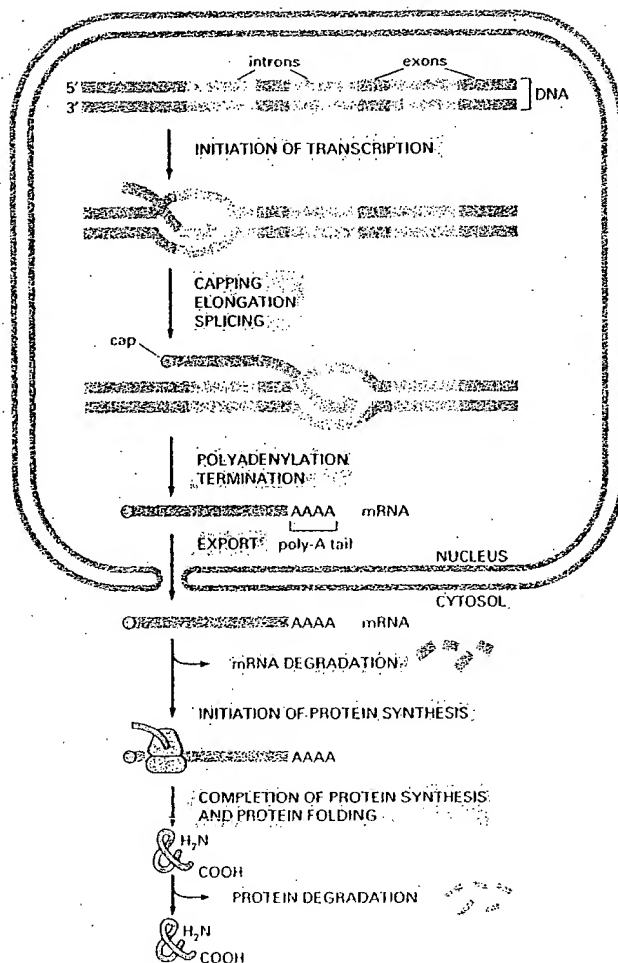


Figure 6-90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6-90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

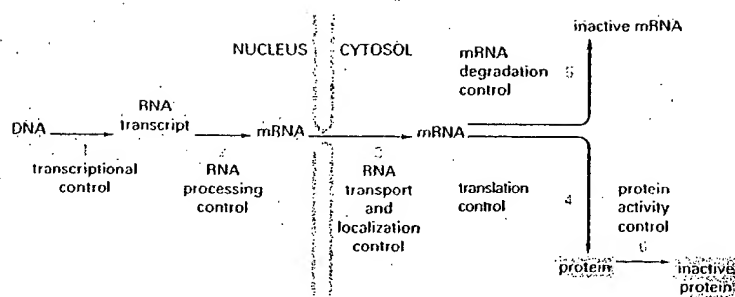


Figure 7-5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7-5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7-5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

266: Mol Cell Biol. 1999 Nov;19(11):7357-68.

Related Articles, Links

FREE full text article at
mcb.asm.org

FREE full text article
in PubMed Central

A sampling of the yeast proteome.

Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JJ.

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.
futcher@cshl.org

In this study, we examined yeast proteins by two-dimensional (2D) gel electrophoresis and gathered quantitative information from about 1,400 spots. We found that there is an enormous range of protein abundance and, for identified spots, a good correlation between protein abundance, mRNA abundance, and codon bias. For each molecule of well-translated mRNA, there were about 4,000 molecules of protein. The relative abundance of proteins was measured in glucose and ethanol media. Protein turnover was examined and found to be insignificant for abundant proteins. Some phosphoproteins were identified. The behavior of proteins in differential centrifugation experiments was examined. Such experiments with 2D gels can give a global view of the yeast proteome.

PMID: 10523624 [PubMed - indexed for MEDLINE]

A Sampling of the Yeast Proteome

B. FUTCHER,^{1*} G. I. LATTER,¹ P. MONARDO,¹ C. S. McLAUGHLIN,² AND J. I. GARRELS³

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724¹; Department of Biological Chemistry, University of California, Irvine, California 92717²; and Proteome, Inc., Beverly, Massachusetts 01915³

Received 15 June 1999/Returned for modification 16 July 1999/Accepted 28 July 1999

In this study, we examined yeast proteins by two-dimensional (2D) gel electrophoresis and gathered quantitative information from about 1,400 spots. We found that there is an enormous range of protein abundance and, for identified spots, a good correlation between protein abundance, mRNA abundance, and codon bias. For each molecule of well-translated mRNA, there were about 4,000 molecules of protein. The relative abundance of proteins was measured in glucose and ethanol media. Protein turnover was examined and found to be insignificant for abundant proteins. Some phosphoproteins were identified. The behavior of proteins in differential centrifugation experiments was examined. Such experiments with 2D gels can give a global view of the yeast proteome.

The sequence of the yeast genome has been determined (9). More recently, the number of mRNA molecules for each expressed gene has been measured (27, 30). The next logical level of analysis is that of the expressed set of proteins. We have begun to analyze the yeast proteome by using two-dimensional (2D) gels.

2D gel electrophoresis separates proteins according to isoelectric point in one dimension and molecular weight in the other dimension (21), allowing resolution of thousands of proteins on a single gel. Although modern imaging and computing techniques can extract quantitative data for each of the spots in a 2D gel, there are only a few cases in which quantitative data have been gathered from 2D gels. 2D gel electrophoresis is almost unique in its ability to examine biological responses over thousands of proteins simultaneously and should therefore allow us a relatively comprehensive view of cellular metabolism.

We and others have worked toward assembling a yeast protein database consisting of a collection of identified spots in 2D gels and of data on each of these spots under various conditions (2, 7, 8, 10, 23, 25). These data could then be used in analyzing a protein or a metabolic process. *Saccharomyces cerevisiae* is a good organism for this approach since it has a well-understood physiology as well as a large number of mutants, and its genome has been sequenced. Given the sequence and the relative lack of introns in *S. cerevisiae*, it is easy to predict the sequence of the primary protein product of most genes. This aids tremendously in identifying these proteins on 2D gels.

There are three pillars on which such a database rests: (i) visualization of many protein spots simultaneously, (ii) quantification of the protein in each spot, and (iii) identification of the gene product for each spot. Our first efforts at visualization and identification for *S. cerevisiae* have been described elsewhere (7, 8). Here we describe quantitative data for these proteins under a variety of experimental conditions.

MATERIALS AND METHODS

Strains and media. *S. cerevisiae* W303 (*MATa ade2-1 his3-11,15 leu2-3, 112 trp1-1 ura3-1 can1-100*) was used (26). -Met YNB (yeast nitrogen base) medium was 1.7 g of YNB (Difco) per liter, 5 g of ammonium sulfate per liter, and

adenine, uracil, and all amino acids except methionine; -Met. -Cys YNB medium was the same but without methionine or cysteine. Medium was supplemented with 2% glucose (for most experiments) or with 2% ethanol (for ethanol experiments). Low-phosphate YEPD was described by Warner (28).

Isotopic labeling of yeast and preparation of cell extracts. Yeast strains were labeled and proteins were extracted as described by Garrels et al. (7, 8). Briefly, cells were grown to 5×10^6 cells per ml at 30°C; 1 ml of culture was transferred to a fresh tube, and 0.3 mCi of [³⁵S]methionine (e.g., Express protein labeling mix; New England Nuclear) was added to this 1-ml culture. The cells were incubated for a further 10 to 15 min and then transferred to a 1.5-ml microcentrifuge tube, chilled on ice, and harvested by centrifugation. The supernatant was removed, and the cell pellet was resuspended in 100 µl of lysis buffer (20 mM Tris-HCl [pH 7.6], 10 mM NaF, 10 mM sodium pyrophosphate, 0.5 mM EDTA, 0.1% deoxycholate; just before use, phenylmethylsulfonyl fluoride was added to 1 mM, leupeptin was added to 1 µg/ml, pepstatin was added to 1 µg/ml, tosyl-sulfonyl phenylalanyl chloromethyl ketone was added to 10 µg/ml, and soybean trypsin inhibitor was added to 10 µg/ml).

The resuspended cells were transferred to a screw-cap 1.5-ml polypropylene tube containing 0.28 g of glass beads (0.5-mm diameter; Biospec Products) or 0.40 g of zirconia beads (0.5-mm diameter; Biospec Products). After the cap was secured, the tube was inserted into a MiniBeadbeater 8 (Biospec Products) and shaken at medium high speed at 4°C for 1 min. Breakage was typically 75%. Tubes were then spun in a microcentrifuge for 10 s at $5,000 \times g$ at 4°C.

With a very fine pipette tip, liquid was withdrawn from the beads and transferred to a prechilled 1.5-ml tube containing 7 µl of DNase I (0.5 mg/ml; Cooper product no. 6330)-RNase A (0.25 mg/ml; Cooper product no. 5679)-Mg (50 mM MgCl₂) mix. Typically 70 µl of liquid was recovered. The mixture was incubated on ice for 10 min to allow the RNase and DNase to work.

Next, 75 µl of 2× SDS (2× SDS is 0.6% sodium dodecyl sulfate [SDS], 2% mercaptoethanol, and 0.1 M Tris-HCl [pH 8]) was added. The tube was plunged into boiling water, incubated for 1 min, and then plunged into ice. After cooling, the tube was centrifuged at 4°C for 3 min at $14,000 \times g$. The supernatant was transferred to a fresh tube and frozen at -70°C. About 5 µl of this supernatant was used for each 2D gel.

2D polyacrylamide gels. 2D gels were made and run as described elsewhere (6-8).

Image analysis of the gels. The Quest II software system was used for quantitative image analysis (20, 22). Two techniques were used to collect quantitative data for analysis by Quest II software. First, before the advent of phosphorimagers, gels were dried and fluorographed. Each gel was exposed to film for three different times (typically 1 day, 2 weeks, and 6 weeks) to increase the dynamic range of the data. The films were scanned along with calibration strips to relate film optical density to disintegrations per minute in the gels and analyzed by the software to obtain a linear relationship between disintegrations per minute in the spots and optical densities of the film images. The quantitative data are expressed as parts per million of the total cellular protein. This value is calculated from the disintegrations per minute of the sample loaded onto the gel and by comparing the film density of each data spot with density of the film over the calibration strips of known radioactivity exposed to the same film. This yields the disintegrations per minute per millimeter for each spot on the gel and thence its parts-per-million value.

After the advent of phosphorimaging, gels bearing ³⁵S-labeled proteins were exposed to phosphorimager screens and scanned by a Fuji phosphorimager, typically for two exposures per gel. Calibration strips of known radioactivity were exposed simultaneously. Scan data from the phosphorimager was assimilated by Quest II software, and quantitative data were recorded for the spots on the gels.

* Corresponding author. Mailing address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724. Phone: (516) 367-8828. Fax: (516) 367-8369. E-mail: fletcher@cshl.org.

Measurements of protein turnover. Cells in exponential phase were pulse-labeled with [35 S]methionine, excess cold Met and Cys were added, and samples of equal volume were taken from the culture at intervals up to 90 min (in one experiment) or up to 160 min (in a second experiment). Incorporation of 35 S into protein was essentially 100% by the first sample (10 min). Extracts were made, and equal fractions of the samples were loaded on 2D gels (i.e., the different samples had different amounts of protein but equal amounts of 35 S). Spots were quantitated with a phosphorimaging and Quest software.

The software was queried for spots whose radioactivity decreased through the time course. The algorithm examined all data points for all spots, drew a best-fit line through the data points, and looked for spots where this line had a statistically significant negative slope. In one of the experiments, there was one such spot. To the eye, this was a minor, unidentified spot seen only in the first two samples (10 and 20 min). In the other experiment, the Quest software found no spots meeting the criteria. Therefore, we concluded that none of the identified spots (and all but one of the visible spots) represented proteins with long half-lives.

Centrifugal fractionation. Cells were labeled, harvested, and broken with glass beads by the standard method described above except that no detergent (i.e., no deoxycholate) was present in the lysis buffer. The crude lysate was cleared of unbroken cells and large debris by centrifugation at $300 \times g$ for 30 s. The supernatant of this centrifugation was then spun at $16,000 \times g$ for 10 min to give the pellet used for Fig. 6B. The supernatant of the $16,000 \times g$, 10-min spin was then spun at $100,000 \times g$ for 30 min to give the supernatant used for Fig. 6A.

Protein abundance calculations. A haploid yeast cell contains about 4×10^{12} g of protein (1, 15). Assuming a mean protein mass of 50 kDa, there are about 50×10^6 molecules of protein per cell. There are about 1.8 methionines per 10 kDa of protein mass, which implies 4.5×10^8 molecules of methionine per cell (neglecting the small pool of free Met). We measured (i) the counts per minute in each spot on the 2D gels, (ii) the total number of counts on each gel (by integrating counts over the entire gel), and (iii) the total number of counts loaded on the gel (by scintillation counting of the original sample). Thus, we know what fraction of the total incorporated radioactivity is present in each spot. After correcting for the methionine (and cysteine [see below]) content of each protein, we calculated an absolute number of protein molecules based on the fraction of radioactivity in each spot and on 50×10^6 total molecules per cell.

The labeling mixture used contained about one-fifth as much radioactive cysteine as radioactive methionine. Therefore, the number of cysteine molecules per protein was also taken into account in calculating the number of molecules of protein, but Cys molecules were weighted one-fifth as heavily as Met molecules.

mRNA abundance calculations. For estimation of mRNA abundance, we used SAGE (serial analysis of gene expression) data (27) and Affymetrix chip hybridization data (29a, 30). The mRNA column in Table 1 shows mRNA abundance calculated from SAGE data alone. However, the SAGE data came from cells growing in YEPD medium, whereas our protein measurements were from cells growing in YNB medium. In addition, SAGE data for low-abundance mRNAs suffers from statistical variation. Therefore, we also used chip hybridization data (29a, 30) for mRNA from cells grown in YNB. These hybridization data also had disadvantages. First, the amounts of high-abundance mRNAs were systematically underestimated, probably because of saturation in the hybridizations, which used 10 μ g of cRNA. For example, the abundance of *ADHI* mRNA was 197 copies per cell by SAGE but only 32 copies per cell by hybridization, and the abundance of *ENO2* mRNA was 248 copies per cell by SAGE but only 41 by hybridization. When the amount of cRNA used in the hybridization was reduced to 1 μ g, the apparent amounts of mRNA were similar to the amounts determined by SAGE (29a, 29b). However, experiments using 1 μ g of cRNA have been done for only some genes (29a). Because amounts of mRNA were normalized to 15,000 per cell, and because the amounts of abundant mRNAs were underestimated, there is a 2.2-fold overestimate of the abundance of nonabundant mRNAs. We calculated this factor of 2.2 by adding together the number of mRNA molecules from a large number of genes expressed at a low level for both SAGE data and hybridization data. The sum for the same genes from hybridization data is 2.2-fold greater than that from SAGE data.

To take into account these difficulties, we compiled a list of "adjusted" mRNA abundance as follows. For all high-abundance mRNAs of our identified proteins, we used SAGE data. For all of these particular mRNAs, chip hybridization suggested that mRNA abundance was the same in YEPD and YNB media. For medium-abundance mRNAs, SAGE data were used, but when hybridization data showed a significant difference between YEPD and YNB, then the SAGE data were adjusted by the appropriate factor. Finally, for low-abundance mRNAs, we used data from chip hybridizations from YNB medium but divided by 2.2 to normalize to the SAGE results. These calculations were completed without reference to protein abundance.

CAI. The codon adaptation index (CAI) was taken from the yeast proteome database (YPD) (13), for which calculations were made according to Sharp and Li (24). Briefly, the index uses a reference set of highly expressed genes to assign a value to each codon, and then a score for a gene is calculated from the frequency of use of the various codons in that gene (24).

Statistical analysis. The JMP program was used with the aid of T. Tully. The JMP program showed that neither mRNA nor protein abundances were normally distributed; therefore, Spearman rank correlation coefficients (r_s) were

calculated. The mRNA (adjusted and unadjusted) and protein data were also transformed so that Pearson product-moment correlation coefficients (r_p) could be calculated. First, this was done by a Box-Cox transformation of log-transformed data. This transformation produced normal distributions, and an r_p of 0.76 was achieved. However, because the Box-Cox transformation is complex, we also did a simpler logarithmic transformation. This produced a normal distribution for the protein data. However, the distribution for the mRNA and adjusted mRNA data was close to, but not quite, normal. Nevertheless, we calculated the r_p and found that it was 0.76, identical to the coefficient from the Box-Cox transformed data. We therefore believe that this correlation coefficient is not misleading, despite the fact that the log(mRNA) distribution is not quite normal.

RESULTS

Visualization of 1,400 spots on three gel systems. Yeast proteins have isoelectric points ranging from 3.1 to 12.8, and masses ranging from less than 10 kDa to 470 kDa. It is difficult to examine all proteins on a single kind of gel, because a gel with the needed range in pI and mass would give poor resolution of the thousands of spots in the central region of the gel. Therefore, we have used three gel systems: (i) pH "4 to 8" with 10% polyacrylamide; (ii) pH "3 to 10" with 10% polyacrylamide; and (iii) nonequilibrium with 15% polyacrylamide (7, 8). Each gel system allows good resolution of a subset of yeast proteins.

Figure 1 shows a pH 4–8, 10% polyacrylamide gel. The pH at the basic end of the isoelectric focusing gel cannot be maintained throughout focusing, and so the proteins resolved on such gels have isoelectric points between pH 4 and pH 6.7. For these pH 4–8 gels, we see 600 to 900 spots on the best gels after multiple exposures.

The pH 3–10 gels (not shown) extend the pI range somewhat beyond pH 7.5, allowing detection of several hundred additional spots. Finally, we use nonequilibrium gels with 15% acrylamide in the second dimension. These allow visualization of about 100 very basic proteins and about 170 small proteins (less than 20 kDa). In total, using all three gel systems, about 1,400 spots can be seen. These represent about 1,200 different proteins, which is about one-quarter to one-third of the proteins expressed under these conditions (27, 30). Here, we focus on the proteins seen on the pH 4–8 gels.

Although nearly all expressed proteins are present on these gels, the number seen is limited by a problem we call coverage. Since there are thousands of proteins on each gel, many proteins comigrate or nearly comigrate. When two proteins are resolved, but are close together, and one protein spot is much more intense than the other, a problem arises in visualizing the weaker spot: at long exposures when the weak signal is strong enough for detection, the signal from the strong spot spreads and covers the signal from the weaker spot. Thus, weak spots can be seen only when they are well separated from strong spots.

For a given gel, the number of detectable spots initially rises with exposure time. However, beyond an optimal exposure, the number of distinguishable spots begins to decrease, because signals from strong spots cover signals from nearby weak spots. At long exposures, the whole autoradiogram turns black. Thus, there is an optimum exposure yielding the maximum number of spots, and at this exposure the weakest spots are not seen.

Largely because of the problem of coverage, the proteins seen are strongly biased toward abundant proteins. All identified proteins have a CAI of 0.18 or more, and we have identified no transcription factors or protein kinases, which are nonabundant proteins. Thus, this technology is useful for examining protein synthesis, amino acid metabolism, and glycolysis but not for examining transcription, DNA replication, or the cell cycle.

Spot identification. The identification of various spots has been described elsewhere (7, 8). At present, 169 different spots representing 148 proteins have been identified. Many of these spots have been independently identified (2, 10, 23, 25). The main methods used in spot identification have been analysis of amino acid composition, gene overexpression, peptide sequencing, and mass spectrometry.

Pulse-chase experiments and protein turnover. Pulse-chase experiments were done to measure protein half-lives (Materials and Methods). Cells were labeled with [35 S]methionine for 10 min, and then an excess of unlabeled methionine was added. Samples were taken at 0, 10, 20, 30, 60, and 90 min after the beginning of the chase. Equal amounts of 35 S were loaded from each sample; 2D gels were run, and spots were quantitated. Surprisingly, almost every spot was nearly constant in amount of radioactivity over the entire time course (not shown). A few spots shifted from one position to another because of post-translational modifications (e.g., phosphorylation of Rpa0 and Efb1). Thus, the proteins being visualized are all or nearly all very stable proteins, with half-lives of more than 90 min. Gygi et al. (10) have come to a similar conclusion by using the N-end rule to predict protein half-lives. This result does not imply that all yeast proteins are stable. The proteins being visualized are abundant proteins; this is partly because they are stable proteins.

Protein quantitation. Because all of the proteins seen had effectively the same half-life, the abundance of each protein was directly proportional to the amount of radioactivity incorporated during labeling. Thus, after taking into account the total number of protein molecules per cell, the average content of methionine and cysteine, and the methionine and cysteine content of each identified protein, we could calculate the abundance of each identified protein (Tables 1 and 2; Materials and Methods). About 1,000 unidentified proteins were also quantified, assuming an average content of Met and Cys.

Many proteins give multiple spots (7, 8). The contribution from each spot was summed to give the total protein amount. However, many proteins probably have minor spots that we are not aware of, causing the amount of protein to be underestimated.

When the proteins on a pH 4–8 gel were ordered by abundance, the most abundant protein had 8,904 ppm, the 10th most abundant had 2,842 ppm, the 100th most abundant had 314 ppm, the 500th most abundant had 57 ppm, and the 1,000th most abundant (visualized at greater than optimum exposure) had 23 ppm. Thus, there is more than a 300-fold range in abundance among the visualized proteins. The most abundant 10 proteins account for about 25% of the total protein on the pH 4–8 gel, the most abundant 60 proteins account for 50%, and the most abundant 500 proteins account for 80%. Since it seems likely that the pH 4–8 gels give a representative sampling of all proteins, we estimate that half of the total cellular protein is accounted for by fewer than 100 different gene products, principally glycolytic enzymes and proteins involved in protein synthesis.

Correlation of protein abundance with mRNA abundance. Estimates of mRNA abundance for each gene have been made by SAGE (27) and by hybridization of cRNA to oligonucleotide arrays (30). These two methods give broadly similar results, yet each method has strengths and weaknesses (Materials and Methods). Table 1 lists the number of molecules of mRNA per cell for each gene studied. One measurement (mRNA) uses data from SAGE analysis alone (27); a second incorporates data from both SAGE and hybridization (30) (adjusted mRNA) (Table 1; Materials and Methods). We correlated protein abundance with mRNA abundance (Fig. 2). For ad-

justed mRNA versus protein, the Spearman rank correlation coefficient, r_s , was 0.74 ($P < 0.0001$), and the Pearson correlation coefficient, r_p , on log transformed data (Materials and Methods) was 0.76 ($P < 0.00001$). We obtained similar correlations for mRNA versus protein and also for other data transformations (Materials and Methods). Thus, several statistical methods show a strong and significant correlation between mRNA abundance and protein abundance. Of course, the correlation is far from perfect; for mRNAs of a given abundance, there is at least a 10-fold range of protein abundance (Fig. 2). Some of this scatter is probably due to posttranscriptional regulation, and some is due to errors in the mRNA or protein data. For example, the protein Yef3 runs poorly on our gels, giving multiple smeared spots. Its abundance has probably been underestimated, partly explaining the low protein/mRNA ratio of Yef3. It is the most extreme outlier in Fig. 2.

These data on mRNA (27, 30) and protein abundance (Table 1) suggest that for each mRNA molecule, there are on average 4,000 molecules of the cognate protein. For instance, for Act1 (actin) there are about 54 molecules of mRNA per cell and about 205,000 molecules of protein. Assuming an mRNA half-life of 30 min (12) and a cell doubling time of 120 min, this suggests that an individual molecule of mRNA might be translated roughly 1,000 times. These calculations are limited to mRNAs for abundant proteins, which are likely to be the mRNAs that are translated best.

A full complement of cell protein is synthesized in about 120 min under these conditions. Thus, 4,000 molecules of protein per molecule of mRNA implies that translation initiates on an mRNA about once every 2 s. This is a remarkably high rate; it implies that if an average mRNA bears 10 ribosomes engaged in translation, then each ribosome completes translation in 20 s; if an average protein has 450 residues; this in turn implies translation of over 20 amino acids per s, a rate considerably higher than estimated for mammals (3 to 8 amino acids per s) (18). These estimates depend on the amount of mRNA per cell (11, 27).

The large number of protein molecules that can be made from a single mRNA raises the issue of how abundance is controlled for less abundant proteins. Many nonabundant proteins may be unstable, and this would reduce the protein/mRNA ratio. In addition, many nonabundant proteins may be translated at suboptimal rates. We have found that mRNAs for nonabundant proteins usually have suboptimal contexts for translational initiation. For example, there are over 600 yeast genes which probably have short open reading frames in the mRNA upstream of the main open reading frame (17a). These may be devices for reducing the amount of protein made from a molecule of mRNA.

Correlation of codon bias with protein abundance. The mRNAs for highly expressed proteins preferentially use some codons rather than others specifying the same amino acid (14). This preference is called codon bias. The codons preferred are those for which the tRNAs are present in the greatest amounts. Use of these codons may make translation faster or more efficient and may decrease misincorporation. These effects are most important for the cell for abundant proteins, and so codon bias is most extreme for abundant proteins. The effect can be dramatic—highly biased mRNAs may use only 25 of the 61 codons.

We asked whether the correlation of codon bias with abundance continues for medium-abundance proteins. There are various mathematical expressions quantifying codon bias; here, we have used the CAI (24) (Materials and Methods) because it gives a result between 0 and 1. The r_s for CAI versus protein abundance is 0.80 ($P < 0.0001$), similar to the mRNA-protein

TABLE 1. Quantitative data^a

Function	Name	CAI	mRNA	Adjusted mRNA	Protein (Glu) (10 ³)	Protein (Eth) (10 ³)	E/G ratio
Carbohydrate metabolism	Adh1	0.810	197	197	1,230	972	0.79
	Adh2	0.504	0		0	963	>20
	Cit2	0.185	1	2.8	23	288	12
	Eno1	0.870	No <i>Nla</i>		410	974	2.4
	Eno2	0.892	248	248	650	215	0.33
	Fba1	0.868	179	179	640	608	0.95
	Hxk1,2	0.500	13	10.5	62	46	
	Icl1	0.251	0		0	671	>20
	Pdb1	0.342	5	5	41	33	
	Pdc1	0.903	226	226	280	205	0.73
	Pfk1	0.465	5	5	75	53	0.71
	Pgi1	0.681	14	14	160	120	0.75
	Pyc1	0.260	1	0.7	37	34	
	Tal1	0.579	5	5	110	35	
	Tdh2	0.904	63	63	430	876	NR
	Tdh3	0.924	460	460	1,670	1,927	NR
	Tpi1	0.817	No <i>Nla</i>		No Met	No Met	
Protein synthesis	Efb1	0.762	33	16.5	358	362	
	Efb1,2	0.801	26	26	99	54	0.55
	Prt1	0.303	4	0.7	12	6	
	Rpa0	0.793	246	246	277	100	0.36
	Tif1,2	0.752	29	29	233	106	0.46
	Yef3	0.777	36	36	14	ND	
Heat shock	Hsc82	0.581	2	2.9	112	75	0.67
	Hsp60	0.381	9	2.3	35	82	2.3
	Hsp82	0.517	2	1.3	52	135	2.6
	Hsp104	0.304	7	7	70	161	2.3
	Kar2	0.439	5	10.1	43	102	2.4
	Ssa1	0.709	2	4.3	303	421	1.4
	Ssa2	0.802	10	5	213	324	1.5
	Ssb1,2	0.850	50	50	270	85	
	Ssc1	0.521	2	2.6	68	80	1.2
	Sse1	0.521	8	8	96	48	
	Sti1	0.247	1	1.1	25	44	1.7
Amino acid synthesis	Ade1	0.229	4	4	14	27	
	Ade3	0.276	2	1.7	12	9	
	Ade5,7	0.257	2	1.4	14	4	
	Arg4	0.229	1	8.1	41	41	
	Gdh1	0.585	10	27	148	55	
	Gln1	0.524	11	11	77	104	1.3
	His4	0.267	3	3	15	23	1.5
	Ilv5	0.801	6	6	152	109	0.7
	Lys9	0.332	4	4	32	17	0.52
	Met6	0.657	No <i>Nla</i>	22	190	80	0.42
	Pro2	0.248	3	3	30	12	
	Ser1	0.258	2	1.2	15	8	
	Trp5	0.319	5	5	28	12	
Miscellaneous	Act1	0.710	54	54	205	164	0.78
	Adk1	0.531	No <i>Nla</i>		47	43	
	Ald6	0.520	3	3	181	159	
	Atp2	0.424	1	4.1	76	109	1.4
	Bmh1	0.322	46	46	191	137	0.72
	Bmh2	0.384	1	1.4	134	147	
	Cdc48	0.306	2	2.4	32	26	
	Cdc60	0.299	2	0.86	6	2	
	Erg20	0.373	5	5	92	39	
	Gpp1	0.603	16	5	234	158	
	Gsp1	0.621	3	3	115	39	
	Ipp1	0.620	4	4	254	147	0.34
	Lcb1	0.173	0.3	0.8	19	40	0.58
	Mol1	0.423	0	0.45	20	16	
	Pab1	0.488	3	3	41	19	0.47
	Psa1	0.600	15	15	148	56	
	Rnr4	0.497	6	6	44	37	
	Sam1	0.494	5	5	59	21	
	Sam2	0.497	3	15	63	20	
	Sod1	0.376	36	36	631	618	
	Uba1	0.212	2	2	14	20	
	YKL056	0.731	62	62	253	112	0.44
	YLR109	0.549	21	21	930		
	YMR116	0.777	41	41	184	40	0.20

^a CAI, a measure of codon bias, is taken from the YPD. mRNA, number of mRNA molecules per cell from SAGE data (27); adjusted mRNA, number of mRNA molecules per cell based on both SAGE and chip hybridization (30) (see Materials and Methods); Protein (Glu), number of molecules of protein per cell in YNB-glucose; Protein (Eth), number of molecules of protein per cell in YNB-ethanol; E/G ratio, ratio of protein abundance in ethanol to glucose. The E/G ratio is not given if it was close to 1 or if it was not repeatable (NR) in multiple gels. Some gene products (e.g., Tif1 and Tif2 [Tif1,2]) were difficult to distinguish on either a protein or an mRNA basis; these are pooled. No *Nla*, there was no suitable *Nla*III site in the 3' region of the gene, and so there are no SAGE mRNA data; No Met, the mature gene product contains no methionines; and so there are no reliable protein data.

TABLE 2. Functions of proteins listed in Table 1

Name ^a	YPD title lines ^b
Adh1	Alcohol dehydrogenase I; cytoplasmic isozyme reducing acetaldehyde to ethanol, regenerating NAD ⁺
Adh2	Alcohol dehydrogenase II; oxidizes ethanol to acetaldehyde, glucose repressed
Cit2	Citrate synthase, peroxisomal (nonmitochondrial); converts acetyl-CoA and oxaloacetate to citrate plus CoA
Eno1	Enolase 1 (2-phosphoglycerate dehydratase); converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Eno2	Enolase 2 (2-phosphoglycerate dehydratase); converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Fba1	Fructose biphosphate aldolase II; sixth step in glycolysis
Hxk1	Hexokinase I; converts hexoses to hexose phosphates in glycolysis; repressed by glucose
Hxk2	Hexokinase II; converts hexoses to hexose phosphates in glycolysis and plays a regulatory role in glucose repression
Icl1	Isocitrate lyase, peroxisomal; carries out part of the glyoxylate cycle; required for gluconeogenesis
Pdb1	Pyruvate dehydrogenase complex, E1 beta subunit
Pdc1	Pyruvate decarboxylase isozyme 1
Pfk1	Phosphofructokinase alpha subunit, part of a complex with Pfk2p which carries out a key regulatory step in glycolysis
Pgi1	Glucose-6-phosphate isomerase, converts glucose-6-phosphate to fructose-6-phosphate
Pyc1	Pyruvate carboxylase 1; converts pyruvate to oxaloacetate for gluconeogenesis
Tal1	Transaldolase; component of nonoxidative part of pentose phosphate pathway
Tdh2	Glyceraldehyde-3-phosphate dehydrogenase 2; converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tdh3	Glyceraldehyde-3-phosphate dehydrogenase 3; converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tpi1	Triosephosphate isomerase; interconverts glyceraldehyde-3-phosphate and dihydroxyacetone phosphate
Efb1	Translation elongation factor EF-1B; GDP/GTP exchange factor for Tef1p/Tef2p
Eft1	Translation elongation factor EF-2; contains diphthamide which is not essential for activity; identical to Eft2p
Eft2	Translation elongation factor EF-2; contains diphthamide which is not essential for activity; identical to Eft1p
Pti1	Translation initiation factor eIF3 beta subunit (p90); has an RNA recognition domain
Rpa0 (RPPO)	Acidic ribosomal protein A0
Tif1	Translation initiation factor 4A (eIF4A) of the DEAD box family
Tif2	Translation initiation factor 4A (eIF4A) of the DEAD box family
Yef3	Translation elongation factor EF-3A; member of ATP-binding cassette superfamily
Hsc82	Chaperonin homologous to <i>E. coli</i> HtpG and mammalian HSP90
Hsp60	Mitochondrial chaperonin that cooperates with Hsp10p; homolog of <i>E. coli</i> GroEL
Hsp82	Heat-inducible chaperonin homologous to <i>E. coli</i> HtpG and mammalian HSP90
Hsp104	Heat shock protein required for induced thermotolerance and for resolubilizing aggregates of denatured proteins; important for [psi ⁻]-to-[psi ⁺] prion conversion
Kar2	Heat shock protein of the endoplasmic reticulum lumen required for protein translocation across the endoplasmic reticulum membrane and for nuclear fusion; member of the HSP70 family
Ssa1	Cytoplasmic chaperone; heat shock protein of the HSP70 family
Ssa2	Cytoplasmic chaperone; member of the HSP70 family
Ssb1	Heat shock protein of HSP70 family involved in the translational apparatus
Ssb2	Heat shock protein of HSP70 family, cytoplasmic
Ssc1	Mitochondrial protein that acts as an import motor with Tim44p and plays a chaperonin role in receiving and folding of protein chains during import; heat shock protein of HSP70 family
Sse1	Heat shock protein of the HSP70 family; multicopy suppressor of mutants with hyperactivated Ras/cyclic AMP pathway
Sui1	Stress-induced protein required for optimal growth at high and low temperature; has tetratricopeptide repeats
Ade1	Phosphoribosylamidoimidazole-succinocarboxamide synthase; catalyzes the seventh step in de novo purine biosynthesis pathway
Ade3	C ₄ tetrahydrofolate synthase (trifunctional enzyme), cytoplasmic
Ade5,7	Phosphoribosylamine-glycine ligase plus phosphoribosylformylglycinamide cyclo-ligase; bifunctional protein
Arg4	Argininosuccinate lyase; catalyzes the final step in arginine biosynthesis
Gdh1	Glutamate dehydrogenase (NADP ⁺); combines ammonia and α-ketoglutarate to form glutamate
Gln1	Glutamine synthetase; combines ammonia to glutamate in ATP-driven reaction
His4	Phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphohydrolase/histidinol dehydrogenase; 2nd, 3rd, and 10th steps of his biosynthesis pathway
Ikv5	Ketol-acid reductoisomerase (acetohydroxy, acid reductoisomerase) (α-keto-β-hydroxylacyl) reductoisomerase; second step in Val and Ile biosynthesis pathway
Lys9	Saccharopine dehydrogenase (NADP ⁺ , L-glutamate forming) (saccharopine reductase), seventh step in lysine biosynthesis pathway
Met6	Homocysteine methyltransferase; (5-methyltetrahydropteroyl triglutamate-homocysteine methyltransferase), methionine synthase, cobalamin independent
Pro2	γ-Glutamyl phosphate reductase (phosphoglutamate dehydrogenase), proline biosynthetic enzyme
Ser1	Phosphoserine transaminase; involved in synthesis of serine from 3-phosphoglycerate
Trp5	Tryptophan synthase, last (5th) step in tryptophan biosynthesis pathway
Act1	Actin; involved in cell polarization, endocytosis, and other cytoskeletal functions
Adk1	Adenylate kinase (GTP:AMP phosphotransferase), cytoplasmic
Ald6	Cytosolic acetaldehyde dehydrogenase
Atp2	Beta subunit of F1-ATP synthase; 3 copies are found in each F1 oligomer
Bmh1	Homolog of mammalian 14-3-3 protein; has strong similarity to Bmh2p
Bmh2	Homolog of mammalian 14-3-3 protein; has strong similarity to Bmh1p
Cdc48	Protein of the AAA family of ATPases; required for cell division and homotypic membrane fusion
Cdc60	Leucyl-tRNA synthetase, cytoplasmic
Erg20	Farnesyl pyrophosphate synthetase; may be rate-limiting step in sterol biosynthesis pathway
Cpp1 (Rhr2)	DL-Glycerol phosphate phosphatase
Gsp1	Ran, a GTP-binding protein of the Ras superfamily involved in trafficking through nuclear pores
Ipp1	Inorganic pyrophosphatase, cytoplasmic
Lcb1	Component of serine C-palmitoyltransferase; first step in biosynthesis of long-chain base component of sphingolipids
Mot1 (Thi4)	Thiamine-repressed protein essential for growth in the absence of thiamine
Pub1	Poly(A)-binding protein of cytoplasm and nucleus; part of the 3'-end RNA-processing complex (cleavage factor I); has 4 RNA recognition domains
Psa1	Mannose-1-phosphate guanylttransferase; GDP-mannose pyrophosphorylase
Rnr4	Ribonucleotide reductase small subunit
Sam1	S-Adenosylmethionine synthetase 1
Sam2	S-Adenosylmethionine synthetase 2
Sod1	Copper-zinc superoxide dismutase
Uba1	Ubiquitin-activating (E1) enzyme
YKL056	Resembles translationally controlled tumor protein of animal cells and higher plants
YLR109 (Ahp1)	Alkyl hydroperoxide reductase
YMR116 (Asc1)	Abundant protein with effects on translational efficiency and cell size, has two WD (WD-40) repeats

^a Accepted name from the *Saccharomyces* genome database and YPD. Names in parentheses represent recent changes.

^b Courtesy of Proteome, Inc., reprinted with permission.

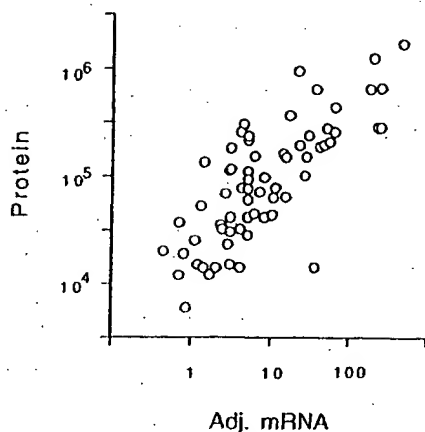


FIG. 2. Correlation of protein abundance with adjusted mRNA abundance. The number of molecules per cell of each protein is plotted against the number of molecules per cell of the cognate mRNA, with an r_p of 0.76. Note the logarithmic axes. Data for mRNA were taken from references 27 and 30 and combined as described in Materials and Methods.

correlation, confirming a strong correlation between CAI and protein abundance (Fig. 3). The relationship between CAI and protein abundance is log linear from about 1,000,000 to about 10,000 molecules per cell. We have no data for rarer proteins.

It is not clear whether CAI reflects maximum or average levels of protein expression. The proteins used for the CAI-protein correlation included some proteins which were not expressed at maximum levels under the condition of the experiment (Hsc82, Hsp104, Ssa1, Ade1, Arg4, His4, and others). When these proteins were removed from consideration and the correlation between CAI and the remaining (presumably constitutive) proteins was recalculated, the r_s was essentially unchanged (not shown).

The equation describing the graph in Fig. 3 is $\log(\text{protein molecules/cell}) = (2.3 \times \text{CAI}) + 3.7$. Thus, under certain conditions (a CAI of 0.3 or greater; a constitutively expressed gene), a very rough estimate of protein abundance can be made by raising 10 to the power of $[(2.3 \times \text{CAI}) + 3.7]$.

The distribution of CAI over the genome (Fig. 4) consists of a lower, bell-shaped distribution, possibly indicating a region where there is no selection for codon bias, and an upper, flat distribution, starting at a CAI of about 0.3, possibly indicating a region where there is selection for codon bias. Almost all of the proteins whose abundance we have measured are in the upper, flat portion of the distribution. In the lower, bell-shaped region, we do not know whether there is a correlation between CAI and protein abundance.

Changes in protein abundance in glucose and ethanol. A comparison of cells grown in glucose (Fig. 1A) with cells grown in ethanol (Fig. 1B) is shown in Table 1. As is well known, some proteins are induced tremendously during growth on ethanol. Two striking examples are the peroxisomal enzymes Icl1 (isocitrate lyase) and Cit2 (citrate synthase), which are induced in ethanol by more than 100- and 12-fold, respectively (Fig. 1; Table 1). These enzymes are key components of the glyoxylate shunt, which diverts some acetyl coenzyme A (acetyl-CoA) from the tricarboxylic acid cycle to gluconeogenesis. *S. cerevisiae* requires large amounts of carbohydrate for its cell wall; in ethanol medium, this carbohydrate comes from gluconeogenesis, which depends on the glyoxylate shunt and on the glycolytic pathway running in reverse. The need for

gluconeogenesis also explains why glycolytic enzymes are abundant even in ethanol medium. Thus, 2D gel analysis shows the prominence of the glycolytic and glyoxylate shunt enzymes in cells grown on ethanol, emphasizing that gluconeogenesis, presumably largely for production of the cell wall, is a major metabolic activity under these conditions.

During gluconeogenesis, substrate-product relationships are reversed for the glycolytic enzymes. One might expect that not all glycolytic enzymes would be well adapted to the reverse reaction. Indeed, 2D gels show that in ethanol, Adh2 (alcohol dehydrogenase 2) is strongly induced (16), while its isozyme Adh1 is not greatly affected. Adh1 and Adh2 each interconvert acetaldehyde and ethanol. Adh1 has a relatively high K_m for ethanol (17 mM), while Adh2 has a lower K_m (0.8 mM) (5). Thus, it is thought that Adh1 is specialized for glycolysis (acetaldehyde to ethanol), while Adh2 is specialized for respiration (ethanol to acetaldehyde) (5, 29). Similarly, Eno1 (enolase 1) is induced in ethanol, while its isozyme Eno2 (enolase 2) decreases in abundance (Table 1) (4, 19). Eno1 is inhibited by 2-phosphoglycerate (the glycolytic substrate), while Eno2 is inhibited by phosphoenolpyruvate (the gluconeogenic substrate) (4). Perhaps Eno1 has a lower K_m for phosphoenolpyruvate than does Eno2, though to our knowledge this has not been tested. Thus, the 2D gels distinguish isozymes specialized for growth on glucose (Adh1 and Eno2) from isozymes specialized for ethanol (Adh2 and Eno1).

Many heat shock proteins (e.g., Hsp60, Hsp82, Hsp104, and Kar2) were about twofold more abundant in ethanol medium than in glucose medium. This is consistent with the increased heat resistance of cells grown in ethanol (3).

Enzymes involved in protein synthesis (Eft1, Rpa0, and Tif1) were about twice as abundant in glucose medium as in ethanol medium. This may reflect the higher growth rate of the cells in glucose.

Phosphorylation of proteins. To examine protein phosphorylation, we labeled cells with ^{32}P and ran 2D gels to examine phosphoproteins. About 300 distinct spots, probably representing 150 to 200 proteins, could be seen on pH 4–8 gels (Fig. 5B). We then aligned autoradiograms of three gels, each with a different kind of labeled protein (^{32}P only [Fig. 5B], ^{32}P plus ^{35}S [Fig. 5A], and ^{35}S only [not shown, but see Fig. 1 for example]). In this way, we made provisional identification of

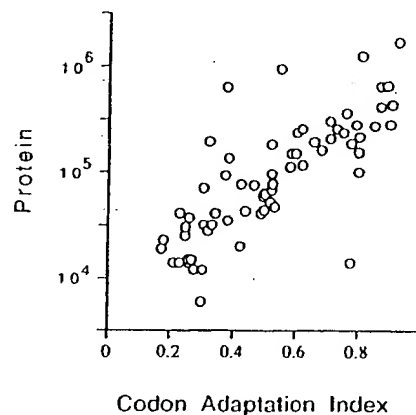


FIG. 3. Correlation of protein abundance with CAI. The number of molecules per cell of each protein is plotted against the CAI for that protein. Note the logarithmic scale on the protein axis. Data for the CAI are from the YPD database (13).

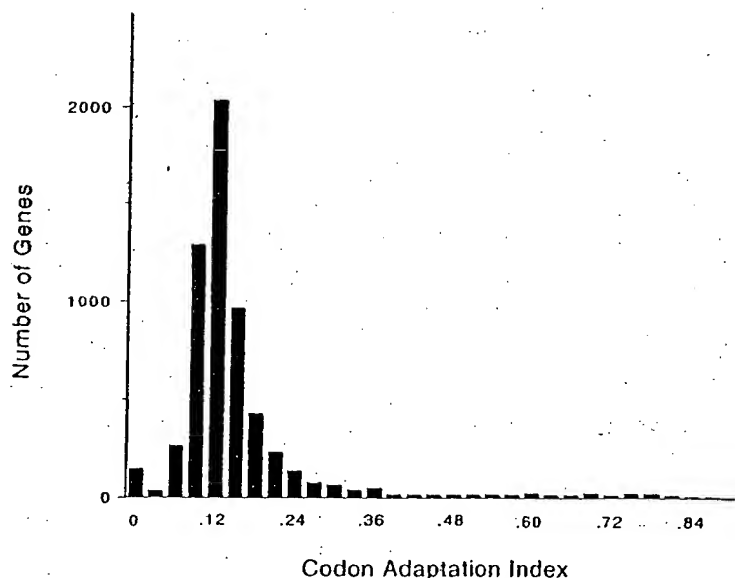


FIG. 4. Distribution of CAI over the whole genome, shown in intervals of 0.030 (i.e., there are 150 genes with a CAI between 0.000 and 0.030, inclusive; 31 genes with a CAI between 0.031 and 0.060; 269 genes with a CAI between 0.061 and 0.090; 1,296 genes with a CAI between 0.091 and 0.120; etc.). The distribution peaks with 2,028 genes with a CAI between 0.121 and 0.150.

some of the ^{32}P -labeled spots as particular ^{35}S -labeled spots. All such identifications are somewhat uncertain, since precise alignments are difficult, and of course multiple spots may exactly comigrate. Nevertheless, we believe that most of the provisional identifications are probably correct. Among the major ^{32}P -labeled proteins are the hexokinases Hxk1 and Hxk2, the acidic ribosome-associated protein Rpa0, the translation factors Yef3 and Efb1, and probably Hsp70 heat shock proteins of the Ssa and Ssb families. Rpa0 and Efb1 are quantitatively monophosphorylated.

Many yeast proteins resolve into multiple spots on these 2D gels (7). Yef3 has five or more spots, at least four of which comigrate with ^{32}P . Tif1 has a major spot showing no ^{32}P labeling and a minor, more acidic spot which overlaps with some ^{32}P label. Tif1 has at least seven spots (7); two of these overlap with some ^{32}P label, but five do not (Fig. 5). Efb1 has at least three spots (7), and none of these overlap with ^{32}P , although there are three nearby, unidentified ^{32}P -labeled spots (a, c, and d in Fig. 5). Spots that seem to be extra forms of Met6, Pdc1, Eno2, and Fba1 can be seen in Fig. 6A, but there is little ^{32}P at these positions in Fig. 5. Thus, phosphorylation explains some but not all of the different protein isoforms seen.

The cell cycle is regulated in part by phosphorylation. We compared ^{32}P -labeled proteins from cells synchronized in G_1 with α -factor, in cells synchronized in G_1 by depletion of G_1 cyclins, and in cells synchronized in M phase with nocodazole. Only very minor differences were seen, and these were difficult to reproduce. The cell cycle proteins regulated by phosphorylation may not be abundant enough for this technique to be applied easily.

Centrifugal fractionation. We fractionated ^{35}S -labeled extracts by centrifugation (Materials and Methods). Figure 6A shows the proteins in the supernatant of a high-speed ($100,000 \times g$, 30 min) centrifugation, while Fig. 6B shows the proteins in the pellet of a low-speed ($16,000 \times g$, 10 min) centrifugation. Many proteins are tremendously enriched in one fraction or the other, while others are present in both.

Most glycolytic enzymes (e.g., Tdh2, Tdh3, Eno2, Pdc1, Adh1, and Fba1) are enriched in the supernatant fraction. The only exception is Pfk1 (not indicated), which is found in both pellet and supernatant fractions. Many proteins involved in protein synthesis (Efb1, Yef3, Prt1, Tif1, and Rpa0) are in the pellet, possibly because of the association of ribosomes with the endoplasmic reticulum. However, Efb1 is in the supernatant, as is a substantial portion of the Efb1. Perhaps surprisingly, several mitochondrial proteins (Atp2 [not shown] and Ilv5) are largely in the supernatant. Perhaps glass bead breakage of cells releases mitochondrial proteins. The nuclear protein Gsp1 is in the pellet fraction. The enrichment produced by centrifugation makes it possible to see minor spots which are otherwise poorly resolved from surrounding proteins. Figure 6B shows that the previously identified Tif1 spot is surrounded by as many as six other spots that cofractionate. We observed six identical or very similar additional spots when we overexpressed Tif1 from a high-copy-number plasmid (not shown). Signal overlaps only one or two of these spots in ^{32}P -labeling experiments (Fig. 5), and so the different forms are not mainly due to different phosphorylation states.

DISCUSSION

Our experience with developing a 2D gel protein database for *S. cerevisiae* is summarized here. With current technology, we can see the most abundant 1,200 proteins, which is about one-third to one-quarter of the proteins expressed. The remaining proteins will be difficult to see and study with the methods that we have used, not because of a lack of sensitivity but because weak spots are covered by nearby strong spots.

Of the 1,200 proteins seen, we have identified 148, with a bias toward the most abundant proteins. Steady application of the methods already used would allow identification of most of the remaining proteins. Gene overexpression will be particularly useful, since it is not affected by the lower abundance of the remaining visible proteins.

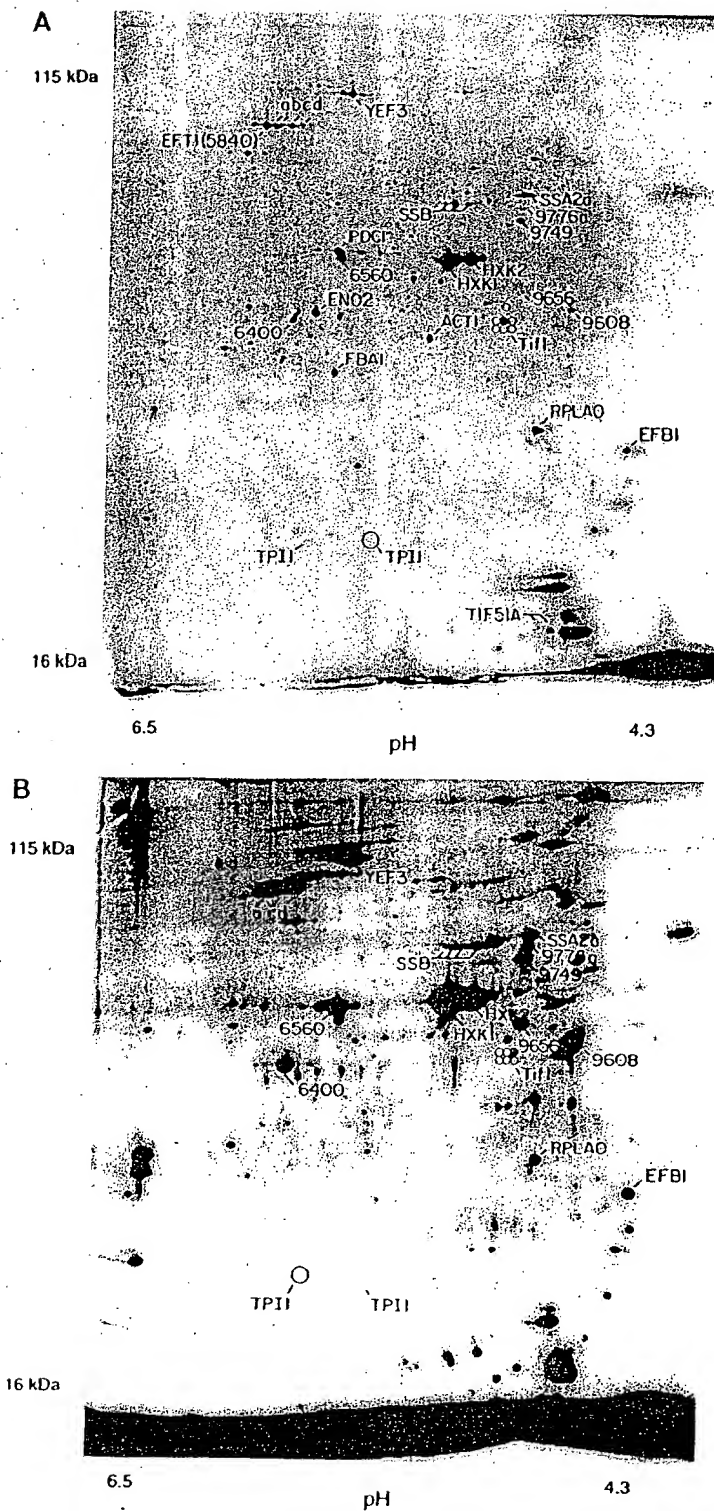


FIG. 5. Phosphorylated proteins. (A) Mixture of ^{32}P -labeled proteins and ^{35}S -labeled proteins. Two separate labeling reactions were done, one with ^{32}P and one with ^{35}S , and extracts were mixed and run on a 2D gel. Spots marked with numbers rather than gene names represent spots noted on ^{35}S gels but unidentified. Spots labeling with ^{32}P were identified by (i) increased labeling compared to the ^{35}S -only gel (not shown); (ii) the characteristic fuzziness of a ^{32}P -labeled spot; and (iii) the decay of signal intensity seen on exposures made 4 weeks later (not shown). A minor form of Tpf1 and at least six minor forms of Tif1 have been noted in overexpression experiments (see also Fig. 6B); positions of the minor forms are indicated by circles. (B) ^{32}P -only labeling. The major form of Tpf1, which is not labeled with ^{32}P , is indicated by a large circle; positions of seven forms of Tif1 are indicated by smaller circles.

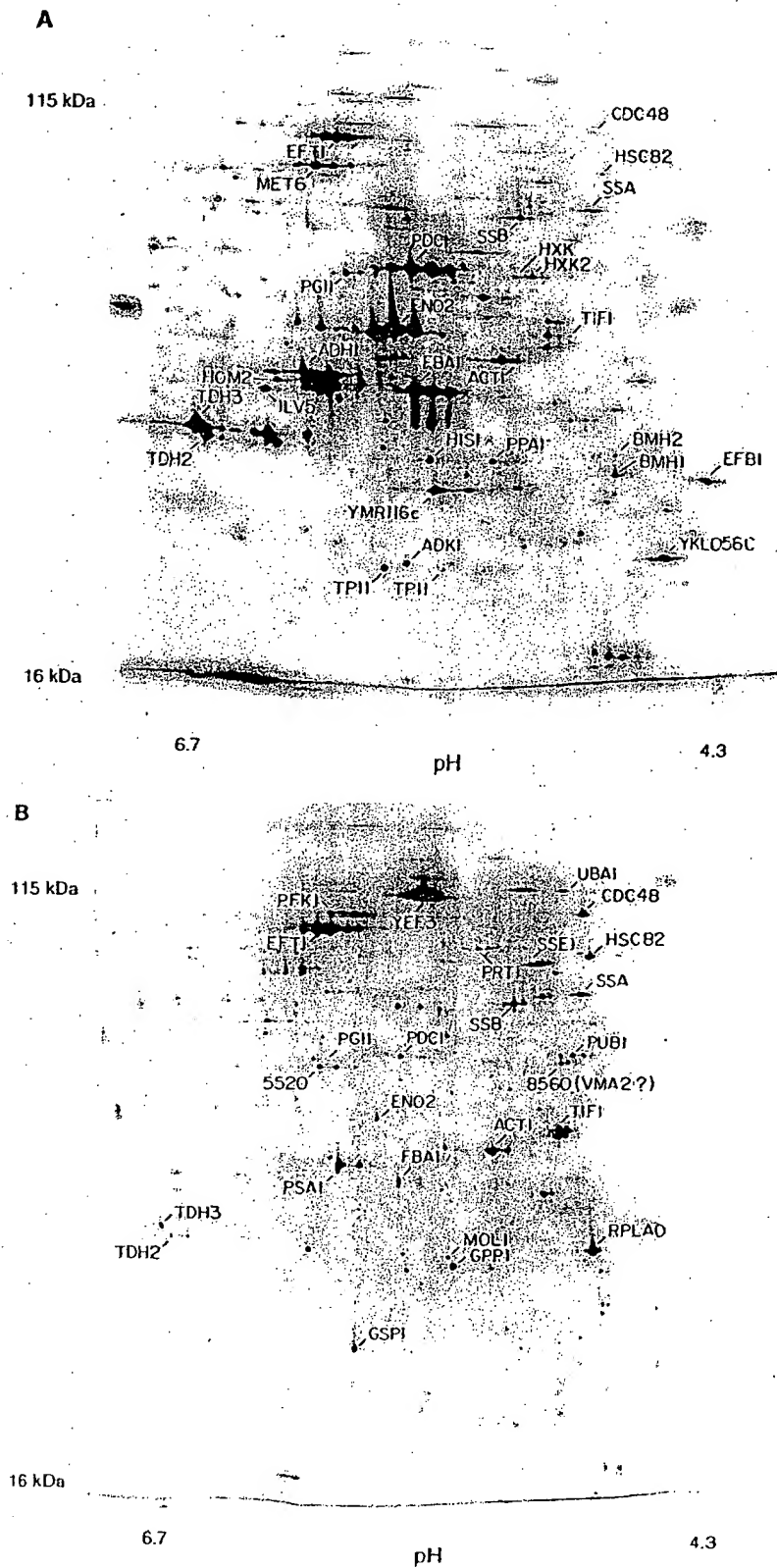


FIG. 6. Fractionation by centrifugation. (A) Proteins in the supernatant of a $100,000 \times g$, 30-min spin; proteins in the pellet of a $16,000 \times g$, 10-min spin. Supernatant fractions examined in multiple experiments done over a wide range of g forces looked similar to each other, as did the pellet fractions.

2D gels of the kind that we have used are not suitable for visualization of rare proteins. However it will be possible to study on a global basis metabolic processes involving relatively abundant proteins, such as protein synthesis, glycolysis, gluconeogenesis, amino acid synthesis, cell wall synthesis, nucleotide synthesis, lipid metabolism, and the heat shock response.

Gygi et al. (10) have recently completed a study similar to ours. Despite generating broadly similar data, Gygi et al. reached markedly different conclusions. We believe that both mRNA abundance and codon bias are useful predictors of protein abundance. However, Gygi et al. feel that mRNA abundance is a poor predictor of protein abundance and that "codon bias is not a predictor of either protein or mRNA levels" (10). These different conclusions are partly a matter of viewpoint. Gygi et al. focus on the fact that the correlations of mRNA and codon bias with protein abundance are far from perfect, while we focus on the fact that, considering the wide range of mRNA and protein abundance and the undoubted presence of other mechanisms affecting protein abundance, the correlations are quite good.

However, the different conclusions are also partly due to different methods of statistical analysis and to real differences in data. With respect to statistics, Gygi et al. used the Pearson product-moment correlation coefficient (r_p) to measure the covariance of mRNA and protein abundance. Depending on the subset of data included, their r_p values ranged from 0.1 to 0.94. Because of the low r_p values with some subsets of the data, Gygi et al. concluded that the correlation of mRNA to protein was poor. However, the r_p correlation is a parametric statistic and so requires variates following a bivariate normal distribution; that is, it would be valid only if both mRNA and protein abundances were normally distributed. In fact, both distributions are very far from normal (data not shown), and so a calculation of r_p is inappropriate. There was no statistical backing for the assertion that codon bias fails to predict protein abundance.

We have taken two statistical approaches. First, we have used the Spearman rank correlation coefficient (r_s). Since this statistic is nonparametric, there is no requirement for the data to be normally distributed. Using the r_s , we find that mRNA abundance is well correlated with protein abundance ($r_s = 0.74$), and the CAI is also well correlated with protein abundance ($r_s = 0.80$) (and also with mRNA abundance [data not shown]). For the data of Gygi et al. (10), we obtained similar results, though with their data the correlation is not as good; $r_s = 0.59$ for the mRNA-to-protein correlation, and $r_s = 0.59$ for the codon bias-to-protein correlation.

In a second approach, we transformed the mRNA and protein data to forms where they were normally distributed, to allow calculation of an r_p (Materials and Methods). Two transformations, Box-Cox and logarithmic, were used; both gave good correlations with our data [e.g., $r_p = 0.76$ for $\log(\text{adjusted RNA})$ to $\log(\text{protein})$]. We were not able to transform the data of Gygi et al. to a normal distribution.

Finally, there are also some differences in data between the two studies. These may be partly due to the different measurement techniques used: Gygi et al. measured protein abundance by cutting spots out of gels and measuring the radioactivity in each spot by scintillation counting, whereas we used phosphorimaging of intact gels coupled to image analysis. We compared our data to theirs for the proteins common between the studies (but excluding proteins whose mRNAs are known to differ between rich and minimal media, and excluding Tif1, which was anomalous in differing by 100-fold between the two data sets). The r_s between the two protein data sets was 0.88 ($P < 0.0001$). Although this is a strong correlation, the fact that

it is less than 1.0 suggests that there may have been errors in measuring protein abundance in one or both studies. After normalizing the two data sets to assume the same amount of protein per cell, we found a systematic tendency for the protein abundance data of Gygi et al. to be slightly higher than ours for the highest-abundance proteins and also for the lowest-abundance proteins but slightly lower than ours for the middle-abundance proteins. These systematic differences suggest some systematic errors in protein measurement. Although we do not know what the errors are, we suggest the following as a reasonable speculation. For the highest-abundance proteins, we may have underestimated the amount of protein because of a slightly nonlinear response of the phosphorimager screens. For the lowest-abundance proteins, Gygi et al. may have overestimated the amount of protein because of difficulties in accurately cutting very small spots out of the gel and because of difficulties in background subtraction for these small, weak spots. The difference in the middle abundance proteins may be a consequence of normalization, given the two errors above.

The low-abundance proteins in the data set of Gygi et al. have a poor correlation with mRNA abundance. We calculate that the r_s is 0.74 for the top 54 proteins of Gygi et al. but only 0.22 for the bottom 53 proteins, a statistically significant difference. However, with our data set, the r_s is 0.62 for the top 33 proteins and 0.56 (not significantly different) for the bottom 33 proteins (which are comparable in abundance to the bottom 53 proteins of Gygi et al.). Thus, our data set maintains a good correlation between mRNA and protein abundance even at low protein abundance. This is consistent with our speculation that protein quantification by phosphorimaging and image analysis may be more accurate for small, weak spots than is cutting out spots followed by scintillation counting. Our relatively good correlations even for nonabundant proteins may also reflect the fact that we used both SAGE data and RNA hybridization data, which is most helpful for the least abundant mRNAs. In summary, we feel that the poor correlation of protein to mRNA for the nonabundant proteins of Gygi et al. may reflect difficulty in accurately measuring these nonabundant proteins and mRNAs, rather than indicating a truly poor correlation *in vivo*. It is not surprising that observed correlations would be poorer with less-abundant proteins and mRNAs, simply because the accuracy of measurement would be worse.

How well can mRNA abundance predict protein abundance? With $r_p = 0.76$ for logarithmically transformed mRNA and protein data, the coefficient of determination, $(r_p)^2$, is 0.58. This means that more than half (in log space) of the variation in protein abundance is explained by variation in mRNA abundance. When converted back to arithmetic values, protein abundances vary over about 200-fold (Table 1), and $(r_p)^2 = 0.58$ for the log data means that of this 200-fold variation, about 20-fold is explained by variation in the abundance of mRNA and about 10-fold is unexplained (but could be due partly to measurement errors). For proteins much less abundant than those considered here, we imagine the *in vivo* correlation between mRNA and protein abundance will be worse, and other regulatory mechanisms such as protein turnover will be more important.

Some important conclusions can be drawn from this sampling of the proteome. First, there is an enormous range of protein abundance, from nearly 2,000,000 molecules per cell for some glycolytic enzymes to about 100 per cell for some cell cycle proteins (26a). Second, about half of all cellular protein is found in fewer than 100 different gene products, which are mostly involved in carbohydrate metabolism or protein synthe-

sis. Third, the correlation between protein abundance and CAI is log linear as far as we can see, which is from about 10,000 protein molecules per cell to about 1,000,000. This is somewhat surprising, because it implies that selective forces for codon bias are significant even at moderate expression levels. It also means that codon bias is a useful predictor of protein abundance even for moderately low bias proteins. Fourth, there is a good correlation between protein abundance and mRNA abundance for the proteins that we have studied. This validates the use of mRNA abundance as a rough predictor of protein abundance, at least for relatively abundant proteins. Fifth, for these abundant proteins, there are about 4,000 molecules of protein for each molecule of mRNA. This last conclusion raises questions as to how the levels of nonabundant proteins are regulated and suggests that protein instability, regulated translation, suboptimal rates of translation, and other mechanisms in addition to transcriptional control may be very important for these proteins.

ACKNOWLEDGMENTS

We thank Neena Sareen and Nick Bizios (CSHL 2D gel laboratory) for production of 2D gels, Tom Volpe for help with some experiments, Corine Driessens for help with calculations and statistics, and Herman Wijnen and Nick Edgington for comments on the manuscript. We especially thank Tim Tully for in-depth statistical analysis and for insightful discussions on statistical interpretations.

This work was supported by grant P41-RR02188 from the NIH Biomedical Research Technology Program, Division of Research Resources, to J.J.G., by Small Business Innovation Research grant R44 GM54110 to Proteome, Inc., by grant DAMD17-94-J4050 from the Army Breast Cancer Program to B.F., and by NIH grant RO1 GM45410 to B.F.

REFERENCES

- Baroni, M. D., E. Martegani, P. Monti, and L. Alberghina. 1989. Cell size modulation by *CDC25* and *RA52* genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9:2715-2723.
- Boucherie, H., F. Sagliocco, R. Joubert, I. Maillet, J. Labarre, and M. Perrot. 1996. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* 17:1683-1699.
- Elliott, B., and B. Futcher. 1993. Stress resistance of yeast cells is largely independent of cell cycle phase. *Yeast* 9:33-42.
- Entian, K. D., B. Meurer, H. Kohler, K. H. Mann, and D. Mecke. 1987. Studies on the regulation of enolases and compartmentation of cytosolic enzymes in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 923:214-221.
- Ganzhorn, A. J., D. W. Green, A. D. Hershey, R. M. Gould, and B. V. Plapp. 1987. Kinetic characterization of yeast alcohol dehydrogenases. Amino acid residue 294 and substrate specificity. *J. Biol. Chem.* 262:3754-3761.
- Garrels, J. I. 1989. The Quest system for quantitative analysis of two-dimensional gels. *J. Biol. Chem.* 264:5269-5282.
- Garrels, J. I., B. Futcher, R. Kobayashi, G. I. Latter, B. Schwender, T. Volpe, J. R. Warner, and C. S. McLaughlin. 1994. Protein identifications for a *Saccharomyces cerevisiae* protein database. *Electrophoresis* 15:1466-1486.
- Garrels, J. I., C. S. McLaughlin, J. R. Warner, B. Futcher, G. I. Latter, R. Kobayashi, B. Schwender, T. Volpe, D. S. Anderson, R. Mesquita-Fuentes, and W. E. Payne. 1997. Proteome studies of *S. cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* 18:1347-1360.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoehsel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. Life with 6000 genes. *Science* 274:563-567.
- Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19:1720-1730.
- Hereford, L. M., and M. Rosbash. 1977. Number and distribution of polyadenylated RNA sequences in yeast. *Cell* 10:453-462.
- Herrick, D., R. Parker, and A. Jacobson. 1990. Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 10:2269-2284.
- Hodges, P. E., A. H. McKee, B. P. Davis, W. E. Payne, and J. I. Garrels. 1999. The Yeast Proteome Database (YPD): a model for the organization of genome-wide functional data. *Nucleic Acids Res.* 27:69-73.
- Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2:13-34.
- Johnston, G. C., F. R. Pringle, and L. H. Hartwell. 1977. Coordination of growth with cell division in the yeast *S. cerevisiae*. *Exp. Cell Res.* 105:79-98.
- Johnston, M., and M. Carlson. 1992. Regulation of carbon and phosphate utilization, p. 193-281. In E. Jones, J. Pringle, and J. Broach (ed.), *The molecular and cellular biology of the yeast Saccharomyces*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Kornblatt, M. J., and A. Klugerman. 1989. Characterization of the enolase isozymes of rabbit brain: kinetic differences between mammalian and yeast enolases. *Biochem. Cell. Biol.* 67:103-107.
- Latter, G., and B. Futcher. Unpublished data.
- Mathews, B., N. Sonenberg, and J. W. B. Hershey. 1996. Origins and targets of translational control, p. 1-29. In J. W. B. Hershey, M. B. Mathews, and N. Sonenberg (ed.), *Translational control*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- McAlister, L., and M. J. Holland. 1982. Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. *J. Biol. Chem.* 257:7181-7188.
- Monardo, P. J., T. Boutell, J. I. Garrels, and G. I. Latter. 1994. A distributed system for two-dimensional gel analysis. *Comput. Appl. Biosci.* 10:137-143.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250:4007-4021.
- Patterson, S. D., and G. I. Latter. 1993. Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques* 15:1076-1083.
- Sagliocco, F., J. C. Guillemot, C. Monribot, J. Capdevielle, M. Perrot, E. Ferrara, and H. Boucherie. 1996. Identification of proteins of the yeast protein map using genetically manipulated strains and peptide-mass fingerprinting. *Yeast* 12:1519-1533.
- Sharp, P. M., and W. H. Li. 1987. The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:281-295.
- Shevchenko, A., O. N. Jensen, A. V. Podilejnikov, F. Sagliocco, M. Wilm, O. Yorn, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 93:14440-14445.
- Thomas, B. J., and R. Rothstein. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56:619-630.
- Tyres, M., and B. Futcher. Unpublished data.
- Velculescu, V. E., L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler. 1997. Characterization of the yeast transcriptome. *Cell* 88:243-251.
- Warner, J. 1991. Labeling of RNA and phosphoproteins in *S. cerevisiae*. *Methods Enzymol.* 194:423-428.
- Wills, C. 1976. Production of yeast alcohol dehydrogenase isoenzymes by selection. *Nature* 261:26-29.
- Wodicka, L. Personal communication.
- Wodicka, L. Unpublished data.
- Wodicka, L., H. Dong, M. Mittmann, M.-H. Jio, and D. J. Lockhart. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15:1359-1367.

Galanin in pituitary adenomas.

Grenback E, Bjellerup P, Wallerman E, Lundblad L, Anggard A, Ericson K, Aman K, Landry M, Schmidt WE, Hokfelt T, Hulting AL.

Department of Molecular Medicine, Endocrine and Diabetes Unit, Karolinska Hospital, S-17176 Stockholm, Sweden. Eva.Grenback@ks.se

Tumor galanin content was measured in extracts from human pituitary adenomas using a specific RIA method for monitoring human galanin. Twenty-two out of twenty-four tumors contained galanin with notably high levels in corticotroph adenomas, varying levels in clinically inactive tumors, and low levels in GH secreting adenomas. Tumor galanin and ACTH contents were closely correlated in all tumors. In four young patients with microadenomas and highly active Mb Cushing tumor galanin was inversely related to tumor volume. The molecular form of tumor galanin, studied with reverse-phase HPLC, was homogeneous with the majority of tumor galanin coeluting with standard human galanin. In the tumors analysed with in situ hybridization there was a good correlation between galanin peptide levels and galanin mRNA expression. In some tumors galanin mRNA and POMC levels coexisted, in others they were essentially in different cell populations. Levels of plasma galanin-LI were not related to tumor galanin concentration, and galanin levels were in the same range in sinus petrosus close to the pituitary venous drainage as in peripheral blood. Corticotrophin releasing hormone injections in two patients caused ACTH, but no detectable galanin release into sinus petrosus. Our results demonstrate that corticotroph, but not GH adenomas, express high levels of galanin, in addition to ACTH, and that in some tumors both polypeptides are synthesised in the same cell population. However, galanin levels in plasma were not influenced by the tumor galanin content.

PMID: 14700749 [PubMed - indexed for MEDLINE]

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Regulatory Peptides 117 (2004) 127–139

REGULATORY
PEPTIDESwww.elsevier.com/locate/regpep

Galanin in pituitary adenomas

Eva Grenbäck^{a,*}, Per Bjellerup^b, Ella Wallerman^a, Lars Lundblad^c,
Anders Änggård^c, Kaj Ericson^d, Katarina Åman^e, Marc Landry^f,
Wolfgang E. Schmidt^g, Tomas Hökfelt^e, Anna-Lena Hulting^a

^aDepartment of Molecular Medicine, Endocrine and Diabetes Unit, Karolinska Hospital, Stockholm S-17176, Sweden

^bDepartment of Clinical Chemistry, Stockholm, Sweden

^cDepartment of Otorhinolaryngology, Stockholm, Sweden

^dDepartment of Neuroradiology, Karolinska Hospital, Stockholm, Sweden

^eDepartment of Neuroscience, Karolinska Institutet, Stockholm, Sweden

^fLaboratoire de Biologie Cellulaire, Université de Bordeaux II, Bordeaux, France

^gDepartment of Medicine I, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany

Received 25 March 2003; received in revised form 23 September 2003; accepted 1 October 2003

Abstract

Tumor galanin content was measured in extracts from human pituitary adenomas using a specific RIA method for monitoring human galanin. Twenty-two out of twenty-four tumors contained galanin with notably high levels in corticotroph adenomas, varying levels in clinically inactive tumors, and low levels in GH secreting adenomas. Tumor galanin and ACTH contents were closely correlated in all tumors. In four young patients with microadenomas and highly active Mb Cushing tumor galanin was inversely related to tumor volume. The molecular form of tumor galanin, studied with reverse-phase HPLC, was homogenous with the majority of tumor galanin coeluting with standard human galanin. In the tumors analysed with *in situ* hybridization there was a good correlation between galanin peptide levels and galanin mRNA expression. In some tumors galanin mRNA and POMC levels coexisted, in others they were essentially in different cell populations. Levels of plasma galanin-LI were not related to tumor galanin concentration, and galanin levels were in the same range in sinus petrosus close to the pituitary venous drainage as in peripheral blood. Corticotrophin releasing hormone injections in two patients caused ACTH, but no detectable galanin release into sinus petrosus. Our results demonstrate that corticotroph, but not GH adenomas, express high levels of galanin, in addition to ACTH, and that in some tumors both polypeptides are synthesised in the same cell population. However, galanin levels in plasma were not influenced by the tumor galanin content.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Acromegaly; ACTH; Cushing's disease; GH; Pituitary tumor; Prolactinoma; Pro-opiomelanocortin

1. Introduction

Galanin was first isolated from porcine intestine in 1983 by Tatemoto et al. [1]. In most species galanin is a 29 amino acid, C-terminally amidated peptide, whereas the human form contains 30 amino acids with a non-amidated C-terminal [2–5]. A major breakthrough was the cloning of the rat galanin gene from the rat anterior pituitary and hypothalamus, and the demonstration of its estrogen sensi-

tivity [6,7]. Galanin has been shown to have multiple biological effects. Thus, it seems to be involved in, for example, regulation of feeding (see Refs. [8–10]), in learning (see Refs. [11,12]), and in responses to nerve injury and pain (see Ref. [13]). The first galanin receptor was cloned by Habert-Ortholi et al. [14] and subsequently two further galanin receptor subtypes were identified [15,16]. The creation of a galanin knock-out mouse [17] and of galanin overexpressing mice [18–20] have provided new insights into the physiological roles of galanin.

Galanin-like immunoreactivity (LI) is spread throughout the central nervous system, including the hypothalamus, with the most intense immunoreactivity in the median eminence [21–24], suggesting participation in neuroendocrine processes controlling anterior pituitary hormone se-

Abbreviations: LI, like immunoreactivity; PAD, pathological anatomical diagnosis; TdT, terminal deoxynucleotidyl transferase.

* Corresponding author. Tel.: +46-8-517-73065; fax: +46-8-517-73096.

E-mail address: Eva.Grenback@ks.se (E. Grenbäck).

0167-0115/\$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.regpep.2003.10.022

cretion (see Refs. [9,25–27]). For example, galanin increases GH secretion not only in animal models but also in humans. Thus, given intravenously (i.v.) to man galanin gives rise to a GH peak, delayed in time but equal to and additive to that produced by GHRH [28,29]. On the other hand, galanin has been shown to inhibit the hypothalamo–pituitary–adrenal (HPA) axis [30–33]. Galanin has been proposed to act as a mitogen in estrogen-induced prolactinomas in the rat [34,35]. Galanin gene expression also shows a direct correlation to somatotroph hyperplasia in GHRH transgenic mice [36].

In the rat pituitary galanin is mainly present in the lactotrophs [7], whereas in the human pituitary galanin is coexpressed with ACTH. Thus, several groups [37–44] have studied galanin-LI in pituitary adenomas establishing its presence in ACTH-producing adenomas, but with varying results as to its presence in somatotroph, lactotroph and clinically inactive adenomas.

The aim of this study was to investigate the presence, concentration and molecular form(s) of galanin in different types of pituitary adenomas using an extraction procedure and a newly developed radioimmunoassay (RIA), as well as reverse-phase (RP)-HPLC. In addition, we wished to compare tumor and serum levels of galanin to some other pituitary hormones, with special reference to the ACTH–cortisol axis. Finally, some tumors were analysed with in situ hybridisation for expression of galanin and POMC. The results of these analyses were correlated to clinical data. Preliminary data from this investigation have previously been reported [45].

2. Material and methods

2.1. Patients

Samples for tumor extraction were obtained from 24 patients with pituitary adenomas (Table 1), 15 women (26–78 years, mean 49.7 years) and 9 men (31–74 years, mean 53.3 years) who underwent transsphenoidal surgery at the Department of Otorhinolaryngology, Karolinska Hospital. Seven patients had hypersecretion of ACTH, six of GH, and one of prolactin (PRL). Ten tumors were clinically inactive (Table 1). The tumors were identified by clinical symptoms and signs as well as magnetic resonance imaging scans. The size of the seven ACTH-producing tumors was measured on contrast-enhanced T1-weighted coronal and sagittal images after correction for scaling. The standard formula for volume measurement of a rotational ellipsoid was used for volume calculations [46]. The size of the remaining tumors was assessed with stage and grade according to Hardy and Wilson (see Ref. [47]). Separate tumor samples were also analysed in the Department of Pathology of the Karolinska Hospital.

The patients with Cushing's disease had clinical symptoms and signs of hypercortisolism. The diagnosis was

confirmed by elevated ACTH and cortisol levels, absent diurnal rhythm, pathological CRH-test ($n=6$) with sinus petrosus sampling ($n=4$) or low and high dose dexamethasone ($n=1$). All had pituitary microadenomas visible on MRI scans. They were all cured clinically and biochemically after surgery. Separate pathological anatomical examination confirmed a pituitary adenoma in six cases. In one case (#6) there was insufficient material for analysis.

Galanin was monitored in two patients during sinus petrosus sampling. Samples were not available for galanin analysis from a further two patients who underwent sinus petrosus sampling, confirming in both cases ACTH producing pituitary adenoma. For control, plasma samples were collected from 22 healthy individuals (12 men, 28–55 years old; 10 women 25–62 years old).

The study was approved by the Ethics Committee of Karolinska Institutet (nos. 96-145 and 99-326).

2.2. Samples

Tumor samples for extraction and in situ hybridisation were collected at surgery, placed on dry ice and kept frozen at -70°C .

Blood samples for galanin measurement were collected in EDTA plasma tubes in the morning fasting state prior to surgery and during sinus petrosus sampling. Samples were placed on ice and centrifuged at 4°C (2400 rpm) for 10 min. Plasma was withdrawn and stored at -70°C until analysis.

2.3. Extraction procedure

For extraction, frozen tumor tissues were weighed, cut, still in frozen state, boiled for 10 min in acetic acid (1 mol/l), cooled on ice, homogenized and centrifuged for 10 min (2000 rpm) at 4°C . After collection of the supernatant, the pellet was resuspended in water, followed by a second boiling, homogenization and centrifugation as above. The supernatants from these two extraction procedures were pooled, lyophilised and stored at 4°C . Samples were reconstituted in phosphate buffer (0.05 mol/l; pH 7.4) before assay.

2.4. Gel permeation chromatography

Gel permeation chromatography was performed on one tumor rich in galanin (#5; Table 1). A Sephadex G-50 Superfine (Pharmacia, Uppsala, Sweden) column (2.6×97 cm) was equilibrated and eluted at room temperature with formic acid (0.1 mol/l) with 0.02% sodium azide. Void volume (V_0) was determined with Dextran blue and total volume (V_t) with ^{22}Na . Human synthetic galanin (Peninsula, Belmont, CA) in RIA buffer was used for column calibration. Tumor extract (50 μl in 1 ml water) was eluted at a flow rate of 15 ml/h, and the eluate was collected in fractions (2 ml), which were lyophilised and stored at 4°C until

Table 1

Twenty-four patients with pituitary tumors operated between 1994 and 1996

Pat. id. #	Tumor type	Age	Sex	T-gal, pmol/mg protein	P-gal, pmol/l	T-ACTH/mg prot	T-ACTH/T-gal
1	ACTH	26	f	620	55	2737.4	4.42
2	ACTH	29	f	587	52	1302.24	2.22
3	ACTH	31	m	111	130	657.27	5.92
4	ACTH	35	m	851	32	72666.67	85.39
5	ACTH	66	m	82.2	-	18860	229.4
6	ACTH	67	f	247	37	4065.04	16.46
7	ACTH	78	f	43.5	38	551.05	12.67
8	GH	29	f	0.04 ^a	60	0.67	16.75
9	GH	40	f	1.87	80	20.39	10.9
10	GH	45	m	0.18	-	7.88	43.7
11	GH	49	f	0.12	54	5.06	42.17
12	GH	51	m	0.22	-	3.07	13.95
13	GII	64	m	0.08	<20	0.37	4.63
14	PRL	44	f	0.13	58	13.03	100.2
15	Inactive	34	f	10.2	116	975.4	95.6
16	Inactive	46	f	199	<20	4261.9	21.4
17	Inactive	52	f	0.16	53	1.23	7.69
18	Inactive	53	m	17.1	44	37.67	2.2
19	Inactive	55	f	99.9	<20	5092.78	50.98
20	Inactive	56	f	0.20	-	0.49	2.45
21	Inactive	62	m	0.21	-	3.48	16.57
22	Inactive	69	f	0.13 ^a	40	6.45	49.61
23	Inactive	72	f	1.85	82	20.16	10.9
24	Inactive	74	m	5.67	38	553.6	97.6

Inactive: clinically inactive adenoma.

Patients are arranged according to tumor type, and age. Sex, tumor galanin content (T-GAL) and plasma galanin (P-GAL) are listed.

^a Below detection limit in the RIA. For the purpose of calculation, they were assigned values of 20 pmol/l, i.e. the limit of detection.

assay. For RIA, each fraction was dissolved in 100 µl RIA buffer.

2.5. Reverse-phase-HPLC

Extracts from the nine adenomas containing the largest amounts of galanin, i.e. seven ACTH-secreting adenomas (#1-7; Table 1) and two clinically inactive adenomas (#16, #19; Table 1) were analysed with RP-HPLC. A 218TP C₁₈-Column (5 µm, 4.6 × 250 nm) from Vydac (Hesperia, CA) was used to separate the galanin-LI. The samples were eluted (1 ml/min) with a linear gradient of

water/acetonitrile (HPLC-grade, Merck, Darmstadt, Germany) with 0.1% trifluoroacetic acid (Merck) increasing from 20% to 70% by 1%/min. Aliquots of 0.5 ml were sampled. The fractions were lyophilized and redissolved in 100 µl RIA buffer before analysis. One sample with very high galanin content (#1; Table 2) was diluted 1:10 before chromatography.

2.6. Radioimmunoassay

Antiserum to human galanin(1-30) was produced as described before [48]. Briefly, human galanin(1-30) was

Table 2

Seven patients with ACTH producing pituitary adenomas

Pat. id. #	T-gal, pmol/mg protein	T-ACTH, pmol/mg protein	P-ACTH, pmol/l	S-Cortisol, nmol/l	U-Cortisol, nmol/24 h
1	620.16	2737.4	60	1190	2360
2	587.31	1302.24	7.3	576	1985
3	111.19	657.27	30	476	3192
4	850.67	72666.67	17	639	3394
5	82.15	18860	36	794	794
6	247.38	4065.04	15	579	1056
7	43.51	551.05	5.2	558	846

The table shows tumor tissue content of galanin (T-GAL) and ACTH (T-ACTH) in pmol/mg protein, morning fasting blood levels of ACTH and cortisol and 24 h urinary cortisol excretion.

Reference ranges: P-ACTH, 2-11 pmol/l.

S-Cortisol, 170-600 nmol/l.

U-Cortisol, 70-500 nmol/24 h.

synthesized by Fmoc-solid phase synthesis and purified to apparent homogeneity by RP-HPLC. Five mg galanin were coupled to 30 mg keyhole limpet hemocyanine with 0.25 mg carbodiimide. New Zealand white rabbits were immunized with 0.25 mg peptide equivalent dissolved in 1 ml saline emulsified in 1.5 ml complete Freund's adjuvant (Sigma, St. Louis, MO). Booster injections at 4-6 weeks intervals were done with 0.08-0.1 mg peptide equivalent mixed with 1.5 ml incomplete Freund's adjuvant. The antiserum used was obtained 10 days after the third booster injection.

Final antibody concentration in the assay was 1:750,000 yielding a 30% binding. Synthetic human galanin (Peninsula) was used as standard and ^{125}I -labelled human galanin (New England Nuclear, NEN, Boston, MA) as tracer. The RIA buffer was a phosphate buffer (0.05 mol/l; pH 7.4) containing 0.1% BSA, 0.02% NaN_3 , 0.01% Triton X, and 10 ml Trasylol/l buffer.

In the assay, 100 μl of standard or samples in duplicates (or in single samples from chromatography) were preincubated with 100 μl antibody for 24 h at 4 °C. After the addition of 100 μl tracer (approx. 8000 cpm) incubation continued for another 24 h. Free and bound radioligand were separated by adding 500 μl sheep anti-rabbit antibodies (Pharmacia decanting suspension; Pharmacia, Uppsala, Sweden). After 30 min incubation and addition of 500 μl water, the samples were centrifuged at 4 °C (3000 rpm) for 15 min and decanted, and the radioactivity in the bound fraction was counted in a gamma counter.

Serial dilutions of tumor extracts showed good parallelism with the standard curve. Non-specific binding in the RIA was 2%. Detection limit was 20 pmol/L and IC₅₀ 312 pmol/l. Intraassay variation was 5% and interassay variation 14%.

Crossreactivity was measured at 50% binding, or at the highest binding for those peptides not attaining a 50% binding. In this assay, there was 62% cross-reactivity with rat galanin, 36% with galanin(1-16), and 0.8% with pancreastatin, whereas less than 0.1% was found for galanin(20-30), galanin message-associated peptide(1-41) or -(44-49), CRH, GHRH, somatostatin, vasoactive intestinal polypeptide, peptide histidine-isoleucine, calcitonin gene-related peptide or oxytocin.

ACTH in tumor tissue was analysed by an immunoradiometric assay (IRMA) from Brahms Diagnostica (Berlin, Germany) using two monoclonal antibodies recognizing different binding sites on the antigen. The assay was carried out in two steps to prevent antigen excess resulting in falsely low values. The tumor samples were diluted stepwise with the zero-standard.

Protein content was analysed with the Bio Rad Protein Assay (BioRad, Munich, Germany).

2.7. Preparation of probes

Altogether three oligonucleotides (Scandinavian Gene Synthesis, K ping, Sweden) were used in this study for in

situ detection of POMC mRNA ($n=1$) and human galanin mRNA ($n=2$). Oligonucleotides were designed as below from published gene sequences.

(A) Oligonucleotide complementary to POMC mRNA [49]:

(i) 5'GCCCTCCCGTGGACTTGGCTCCGGACGGC-CATCTCCCCCGCCGTCTCTTCTCTC3'

(B) Oligonucleotides complementary to human galanin mRNA [3]:

(i) 5'CTGTGGTTGCCAACGGCATGTGGGCCCA-GCAGGTAGCCCGCGCTGTTTC3'

(ii) 5'GTCAAAGCTTCCTGGTTTCATGTCATC-TTCGGGCCGACGACCCGCTT3'

All oligonucleotides were chosen in regions presenting few homologies with sequences of related mRNAs, and they were checked against the Genbank database.

The oligonucleotides complementary to human galanin mRNA were labelled as previously described [50] at the 3'-end using terminal deoxynucleotidyl transferase (TdT) (Amersham, Amersham, UK) in a cobalt-containing buffer with ^{35}S -dATP (New England Nuclear, Boston, MA, USA) to a specific activity of $1-4 \times 10^9$ cpm/ μg and purified in Nensorb-20 columns (New England Nuclear). The oligonucleotide complementary to POMC mRNA was labelled by tailing at the 3'-end with digoxigenin-11-dUTP (Boehringer Mannheim, Mannheim, Germany) according to published protocols [51]. Briefly, 100 pmol of each probe were incubated in a final volume of 20 μl with 1 nmol of digoxigenin-11-dUTP, 9 nmol of dATP (Sigma), 50 units of TdT and 4.2 μl of cobalt-containing buffer. After 45 min at 37 °C the reaction was stopped before purification by ethanol precipitation, and probes were then stored at -20 °C.

2.8. In situ hybridization procedures

The frozen pituitary tumors were cut at 14 μm thickness in a cryostat (Microm, Heidelberg, Germany) and thaw-mounted onto Probe On slides (Fisher Scientific, Pittsburgh, PA). The sections were then processed as described earlier [50]. In brief, tissue sections were air-dried and incubated for 16 h at 42 °C with 0.5 ng of each of the two radioactively labelled human galanin probes and 10 nmol/l of the digoxigenin-labelled POMC probe. The probes were diluted in a hybridization solution containing 50% deionized formamide (Baker, Deventer, The Netherlands), $4 \times$ standard saline citrate (SSC, $1 \times \text{SSC} = 0.15 \text{ M NaCl}$ and $0.015 \text{ M NaCitrate}$), $1 \times$ Denhardt's solution (0.02% bovine serum albumin, 0.02% Ficoll (Pharmacia), 0.02% polyvinylpyrrolidone), 0.02 M NaPO_4 (pH 7.0), 1% *N*-lauroyl-sarcosine, 10% dextran sulphate (Pharmacia), 500 mg/l denatured salmon testis DNA (Sigma) and 200 mM dithiothreitol (Sigma). After hybridization, the sections were rinsed in $1 \times \text{SSC}$, $4 \times 15 \text{ min}$ at 55 °C followed by 30 min

at room temperature. The sections were then immersed for 30 min in buffer A (0.1 M Tris, pH 7.5, 1 M NaCl, 2 mM $MgCl_2$) containing 0.5% bovine serum albumin (Sigma) and incubated overnight at 4 °C in the same solution with alkaline-phosphatase conjugated anti-digoxigenin F(ab) fragment (1:5000; Boehringer Mannheim). Afterwards, they were rinsed 3×10 min in buffer A and twice in buffer B (0.1 M Tris, pH 9.5, 0.1 M NaCl, 5 mM $MgCl_2$). Alkaline-phosphatase activity was developed by incubating the sections with 44 μ l NBT and 33 μ l BCIP (Gibco, Gaithersburg, MD) diluted in 10 ml buffer B. The enzymatic reaction was stopped by extensive rinsings in distilled water and in buffer B.

The sections were air dried and coated twice with 3% collodion dissolved in amyl acetate (LKB, Stockholm, Sweden) [52]. They were then dipped into Ilford K5 nuclear emulsion (Ilford, Mobberly, Cheshire, UK) diluted 1:1 with distilled water, exposed for 3 weeks, developed in Kodak D19 for 3 min and fixed in Kodak 3000 for 8 min. After mounting in glycerol, the sections were analysed in a Nikon Microphot-FX microscope equipped with a dark field condenser.

2.9. Statistical analysis

Results are presented as mean \pm SEM, unless otherwise stated. Comparisons between groups were assessed by Mann–Whitney Rank Sum test or by one-way ANOVA, followed by Duncan's Multiple Comparison procedure. When applicable, non-normally distributed variables were log transformed before analysis, in order to get a more closely approximated Gaussian distribution. Correlations were performed using least square linear regression analysis. For the purpose of calculation, undetectable levels of galanin (#8, 22) were assigned values of 20 pmol/l. Statistical significance was set at $p < 0.05$. Statistical analyses

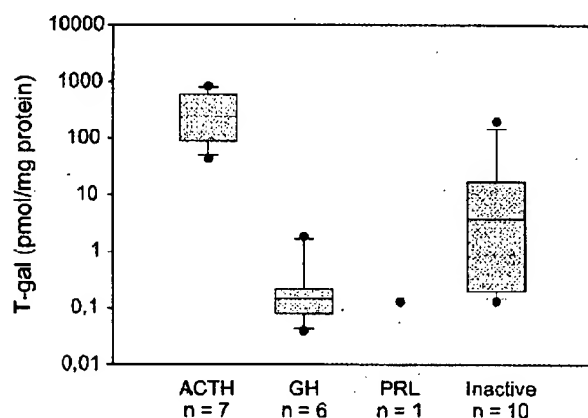


Fig. 1. Galanin concentration in 24 pituitary adenomas arranged according to tumor type. Box plots indicate the median and the lower and upper quartiles, the whiskers the 10th and 90th percentiles, and the outliers are shown as circles.

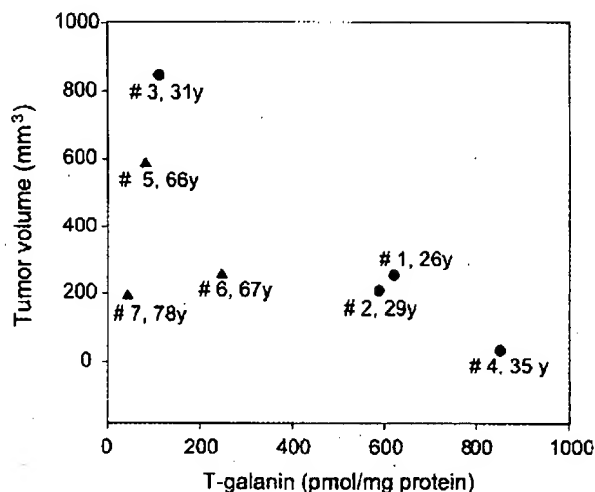


Fig. 2. Tumor galanin content related to tumor volume in seven patients with ACTH secreting adenomas. Patients number and age are given in the figure. Dots indicate young patients, triangles old ones.

were performed using SigmaStat for Windows (Jandel Scientific, Erkrath, Germany).

3. Results

3.1. Tumor samples

Galanin-LI was present in extracts from 22 of the 24 pituitary adenomas examined in this study (Table 1; Fig. 1). High levels of galanin-LI ranging between 43.5 and 851 pmol/mg protein (mean 363 pmol/mg) were found in the seven adenomas from patients with Cushing's disease (#1–7, Table 2). The six GH-producing adenomas contained low

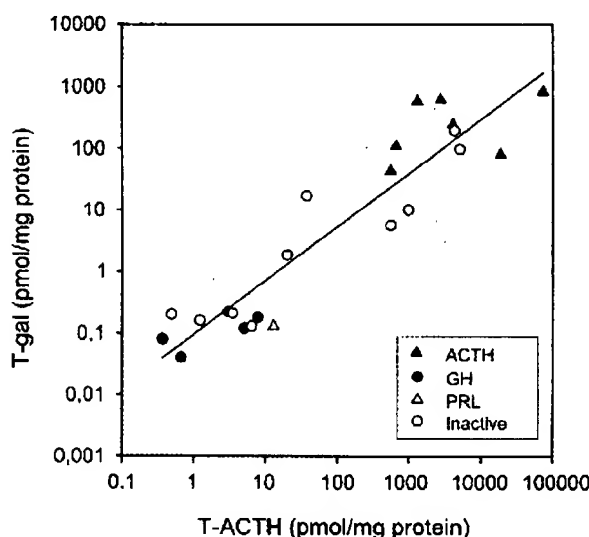


Fig. 3. Tumor galanin content related to tumor ACTH content in 24 pituitary adenomas specified to tumor type. There is a good correlation between galanin and ACTH tumor concentration ($r = 0.932$; $p < 0.001$).

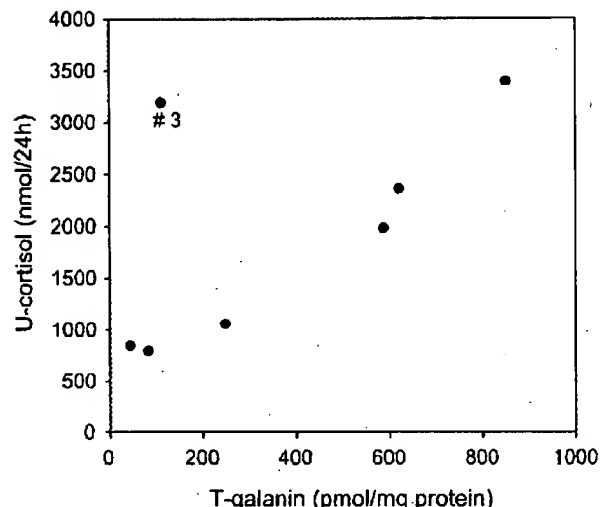


Fig. 4. Tumor galanin content related to 24 h urinary cortisol excretion in seven patients with ACTH secreting pituitary adenomas. Exclusion of patient #3 results in $r=0.974$; $p<0.001$. If included $r=0.614$; $p<0.142$.

amounts of galanin-LI, from 0.04 to 1.87 pmol/mg protein (mean 0.41 pmol/mg) (#8-13, Table 1). Also, the single prolactinoma had very low galanin-LI (#14, Table 1).

The 10 clinically inactive adenomas contained varying amounts of galanin-LI (#15-24, Table 1). Five of these adenomas had low galanin content, ranging from 0.13 to 1.85 pmol/mg (#17, 20-23, Table 1), three had levels in the

intermediate range (5.67-17.1 pmol/mg; #15, 18, 24), whereas two adenomas were rich in galanin-LI (99.9 and 199 pmol/mg; #16, 19, respectively). There was no correlation between galanin levels, or patients age, or sex, irrespective of tumor type. Tumor volume did not correlate to tumor galanin or ACTH-content in the seven patients with Mb Cushing (Fig. 2) ($r=0.638$; $p=0.123$). However, the four youngest patients, who also exhibited the highest cortisol per 24 h (#1-4, Table 2) had an inverse relation between tumor-galanin and volume (Fig. 2).

ACTH was detectable in all adenomas studied, and there was a close correlation between galanin-LI and ACTH levels in the adenomas as a group ($r=0.932$, $p<0.001$; Fig. 3), as well as in the clinically inactive adenomas ($r=0.919$, $p<0.001$), but a less strong correlation among the GII adenomas ($r=0.8429$, $p<0.05$). The seven corticotroph adenomas did not show any correlation between galanin and ACTH content ($r=0.434$). In this group of patients 24-h urinary cortisol excretion correlated strongly to tumor galanin-LI in six out of seven patients (Fig. 4). In the seventh patient (#3) comparatively low tumor galanin-LI and ACTH values were associated with high urinary cortisol. This patient had the largest tumor in the Mb Cushing group. No correlation was found between urinary cortisol excretion and tumor ACTH levels, nor did plasma ACTH or serum cortisol relate to tumor galanin-LI or ACTH content (Table 2). After surgery, all these patients had subnormal cortisol levels. One patient (#2) had recurrent disease a few years later.

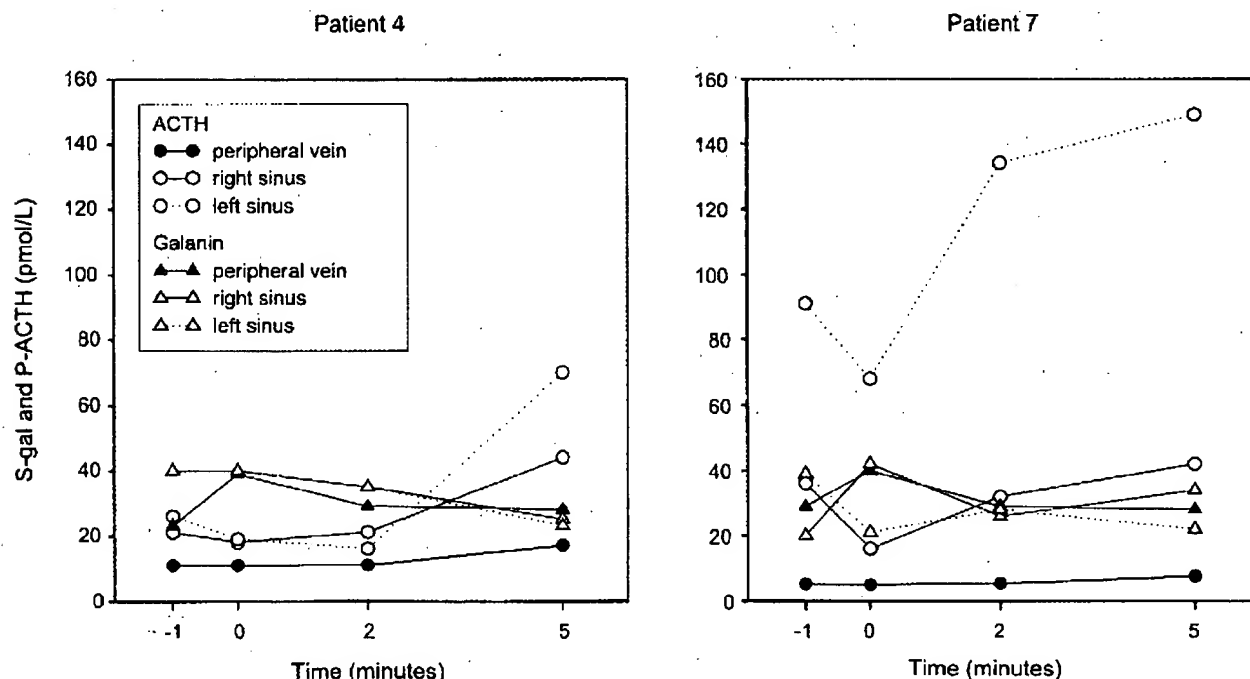


Fig. 5. Sinus petrosus sampling during CRH stimulation test in two patients with ACTH-producing tumors (#4 and 7; Table 1). After injection of 100 µg human CRH (Corticobiss®, Bissendorf, The Netherlands), samples were drawn from right and left cavernous sinus and peripheral vein at -1, 0, 2, and 5 min and analyzed for ACTH and galanin.

3.2. Plasma samples

Galanin-LI in peripheral blood was measured in all but five patients (Table 1). The levels ranged from undetectable to 130 pmol/l (mean 54.2 ± 30 pmol/l). There was, however, no correlation to tumor type, nor to tumor size, galanin-LI content, patient age or sex (not shown).

Galanin-LI in plasma from healthy controls was 93.7 ± 44 pmol/l (101.9–48.7 pmol/l in the men and 83.9–37.6 pmol/l in the women), i.e. it was higher in controls than in the patient group. These control samples were collected 5 years after the patient samples and analysed with a slightly modified RIA including Trasylol to further prevent peptide degradation. Galanin-LI and ACTH were measured in sinus petrosus blood samples from two patients with Cushing's disease (#4, 7) after CRH stimulation (Fig. 5). A rise in plasma ACTH was noted in the left sinus of both patients, with no change in peripheral ACTH levels. No effect was observed on levels of galanin-LI in sinus petrosus or in the periphery, nor did basal plasma concentrations of galanin-LI in sinus petrosus differ from those in the periphery. Thus, there was no correspondence to the high levels of galanin-LI detected in tumor extracts from these two patients.

3.3. Chromatography

Gel permeation chromatography of one tumor (#5, Table 1) revealed one major peak of galanin-LI with the same retention as synthetic human galanin. RP-HPLC of nine adenoma extracts (Fig. 6) showed that in eight of these tumors the majority of galanin-LI had the same retention as synthetic human galanin (16 min) (Fig. 6A). The main peak with galanin-like material was preceded by a smaller one with an elution time of 9–11 min in six extracts and of 13 min in two tissue extracts. One adenoma (#6, Table 1) differed from the rest, in that there was only one major chromatographic peak with galanin-like material with a longer retention time (20 min) coinciding with synthetic galanin(1–16) (Fig. 6B).

3.4. In situ hybridization

The in situ hybridization analysis showed that the pieces of the various tumors processed with this technique could be divided into at least three groups with regard to galanin expression. (i) A galanin signal was found in a large number of cells, sometimes apparently all cells of the tissue piece;

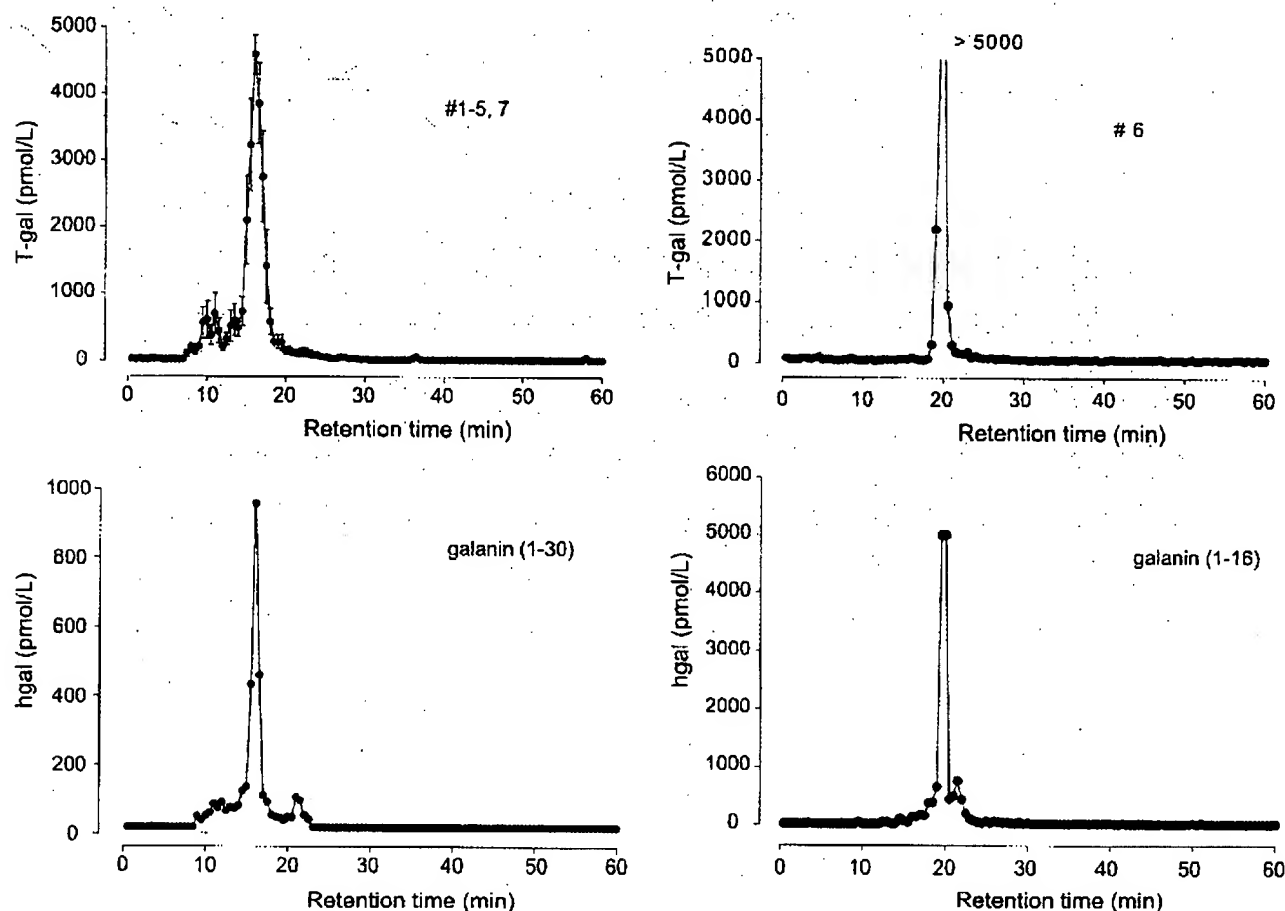


Fig. 6. (A, B) Reverse-phase high performance liquid chromatography of galanin immunoreactivity in pituitary extracts. (A) Results are mean \pm SEM of eight pituitary adenomas. (B) Results of one pituitary adenoma. (C) Elution of human galanin (1–30). (D) Elution of galanin 1–16. Please note different scale in (C).

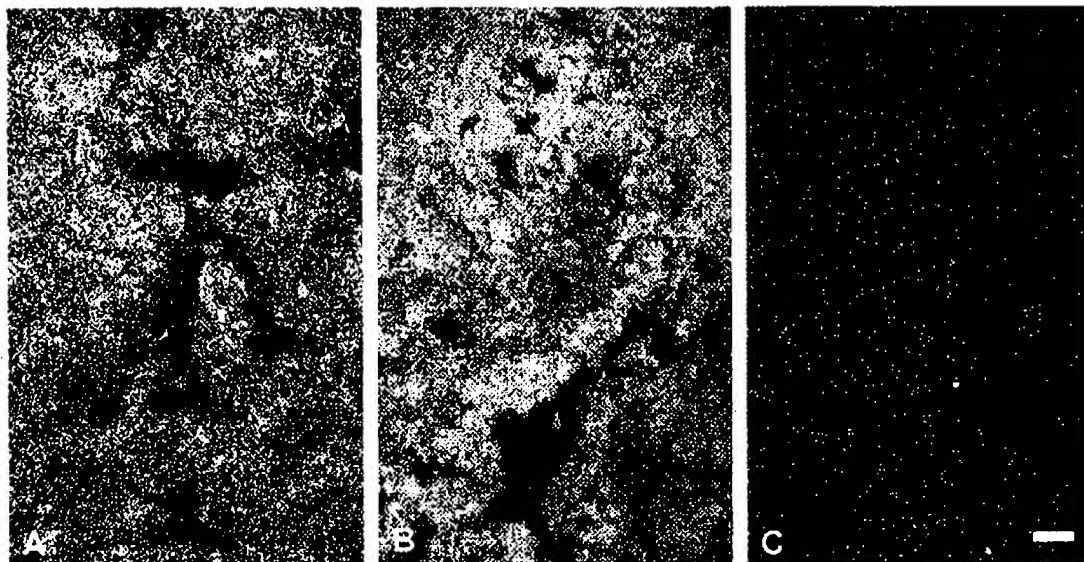


Fig. 7. (A-C). Darkfield micrographs of ACTH-producing adenomas (A, C, #1; B, #2) after hybridization with probes complementary to prepro-galanin mRNA (A, B) and after addition of excess of cold probe (C). A very strong signal is seen evenly distributed over both tumors (A, B) but only few grains in the controls (C). Scale bar indicates 50 μ m for all micrographs.

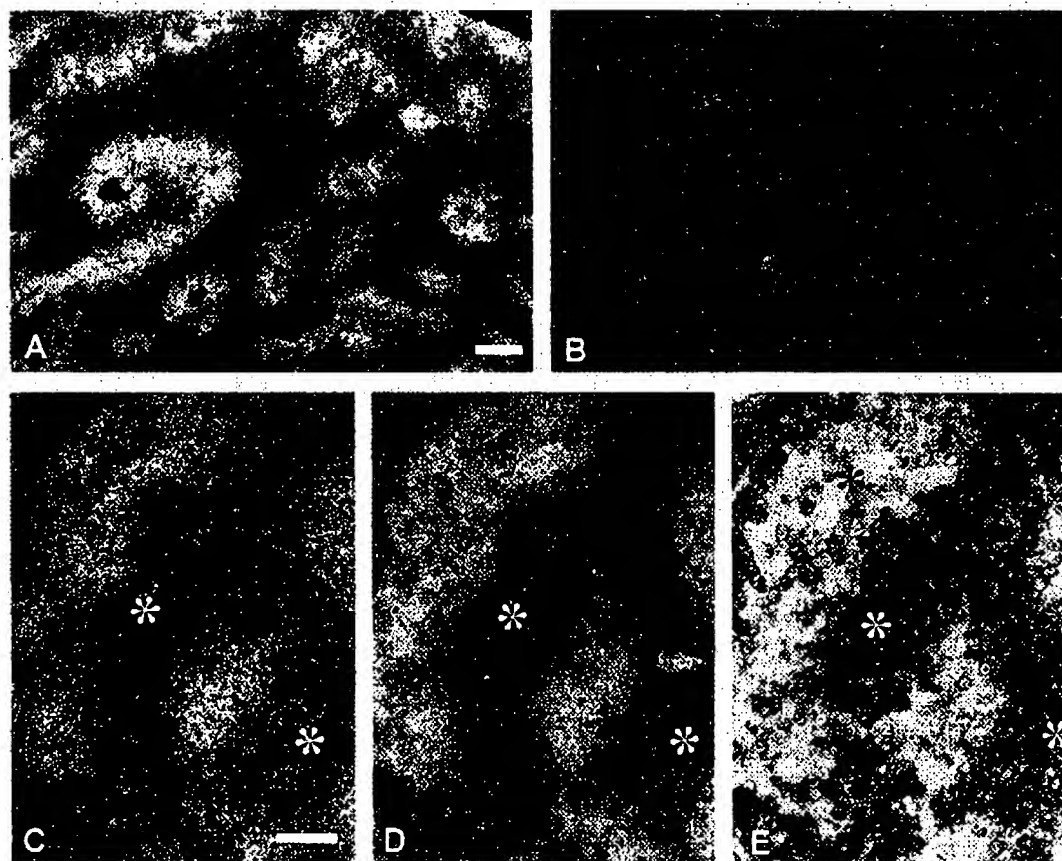


Fig. 8. (A-E) Darkfield (A-C), brightfield (E) and combined dark-and brightfield (D) micrographs of ACTH-producing adenoma (#4) after double-hybridization with radioactive probes complementary to prepro-galanin mRNA (A, C, D) and non-radioactive probe complementary to POMC mRNA (D, E) or after hybridization with an excess of galanin cold probe (B). A strong galanin signal (bright grains) is seen in strands within the tumor tissue (A, C), and they are complementary to the POMC mRNA (dark labelling) distribution (C-E). No signal is seen in the control section (B). Scale bars indicate 50 μ m (A-B; C=D=E).

(ii) a limited number of cells expressed galanin mRNA; and (iii) no galanin mRNA expression could be detected. Two tumors showed an extremely strong signal that could not be blocked by an excess of cold probe, that is the signal was considered to be unspecific.

To group (i), i.e. strongly galanin mRNA expressing tumors, belonged patients #1 (Fig. 7A), #2 (Fig. 7B) and #4 (Fig. 8A, C), whereby one of the two pieces analysed of tumor #2 showed a less compact distribution (data not shown) and #4 showed a markedly uneven distribution with

distinct bands of positive cell groups separated by areas with a low signal (Fig. 8A, C). Tumor #5 showed a signal all over the sections but of a comparatively low intensity (data not shown). To the second group (ii) with a limited number of cells expressing galanin mRNA belonged tumors #3 (Fig. 9A) and #19 (Fig. 9D); in the latter galanin-positive cells often formed small groups (Fig. 9E). In fact, in general the tissue piece of this tumor processed for in situ hybridization appeared to have a normal pituitary morphology at the cellular level (Fig. 9E). In group (iii) no detectable signal

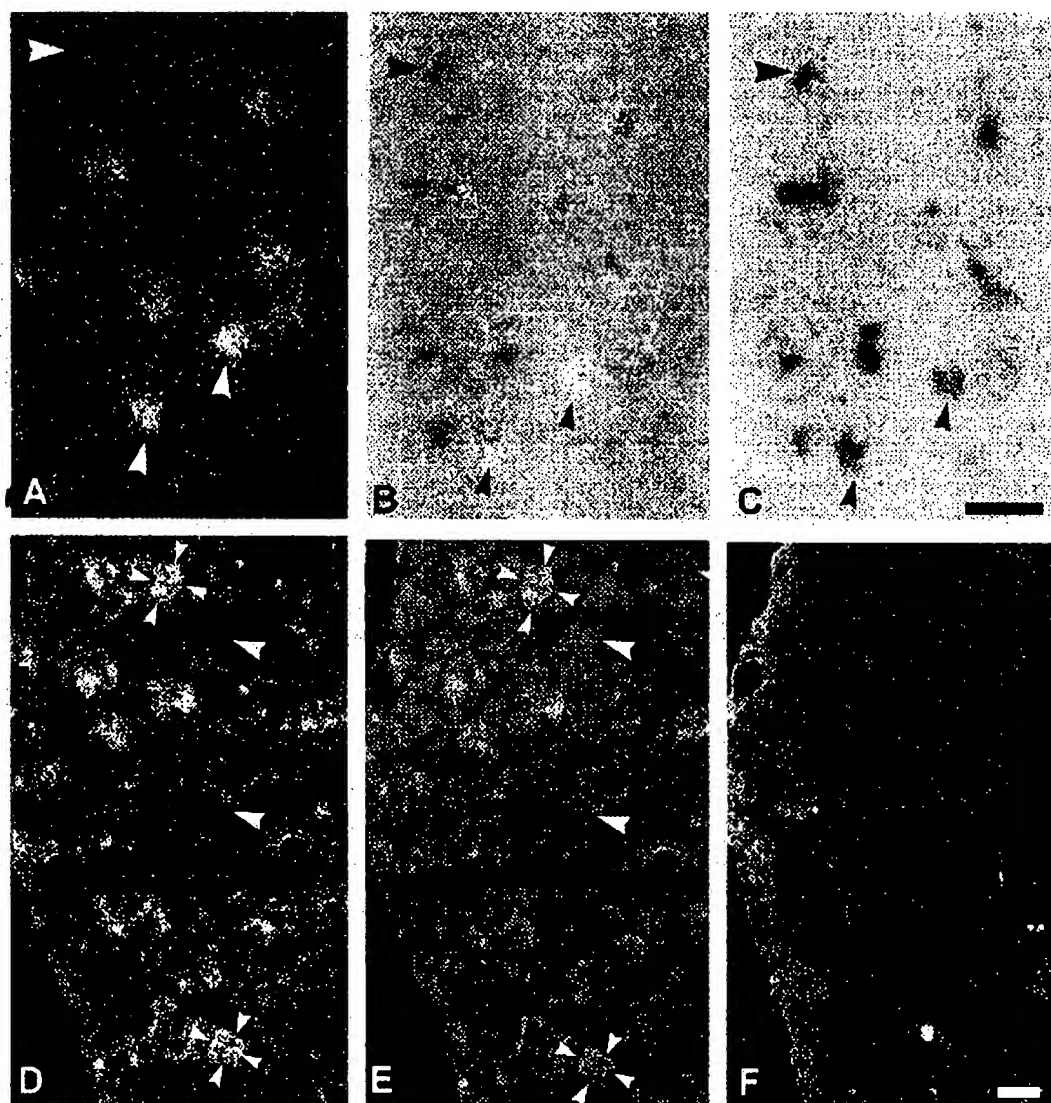


Fig. 9. (A–F) Darkfield (A, D–P), brightfield (C) and combined dark-and brightfield (B) micrographs of ACTH-producing adenoma (#3) (A–C) and of a clinically inactive tumor (#19) (D–F) after double-hybridization with radioactive probe complementary to prepro-galanin mRNA and non-radioactive probe complementary to POMC mRNA (A–C) or after hybridization with probe complementary to prepro-galanin mRNA without (D) or with (E) bisbenzimidazole contrast staining, or after hybridization with an excess of cold probe added to the hybridization cocktail (F). (A–C) Double in situ hybridization reveals several cells expressing both prepro-galanin mRNA (bright grains) and POMC mRNA (dark labelling) (small arrowheads) as well as POMC mRNA-positive and prepro-galanin mRNA-negative cells (big arrowheads). (D–F) Note strong signals in restricted areas of the tissue (A) which in B can often be seen to consist of small groups of cells (small arrowheads). Big arrowheads point to negative cell groups. Note low signal in the control experiment (F). Scale bars indicate 50 μ m (A=B=C; D=E=F).

was observed in tumors #10, #16, #17, or #21 (data not shown).

Some tumors were processed for double in situ hybridization with a non-radioactive POMC mRNA probe. A high degree of coexistence was observed in tumors #1 and #3, whereby this was observed in the majority of cells in tumor #3, but POMC mRNA-positive, galanin mRNA-negative cells were also observed (Fig. 9A-C). In tumor #1 all galanin mRNA positive cells had a POMC mRNA signal. Tumor #4 showed a clear separation between the non-radioactive POMC mRNA signal and the radioactive galanin signal (cf. Fig. 8E with C and D), suggesting synthesis in separate cells. However, POMC mRNA-galanin mRNA coexistence situations could not be excluded. Finally, in some tumors (eg. #10 and #21) a very strong signal was observed after incubation with the galanin mRNA probes. In all other tumors described above (Figs. 7C, 8B and 9F), the signal could be blocked by addition of an excess ($100 \times$) of cold probe, except in tumors #10 and #21. They had a very strong signal which was observed to the same extent after addition of cold probe (data not shown). In agreement, both tumors (#10, 21) had very low galanin levels as monitored with RIA.

4. Discussion

In this study, using a RIA for human galanin, we demonstrate high levels of this peptide in extracts from ACTH-producing pituitary adenomas, moderate and varying galanin levels in clinically non-functioning adenomas and low or not detectable levels in GH- and PRL-producing adenomas, with a good correlation to ACTH levels in the adenomas. Moreover, the RIA data were supported by the in situ hybridization results; a significant galanin signal was only found in the corticotroph adenomas. Double in situ hybridization revealed coexistence of galanin mRNA and POMC mRNA in most tumors, although a clear cellular dissociation of the two transcripts was seen in some other tumours.

When discussing these results, one should keep in mind that different parts of the tumor tissue sample were analyzed in each of the various assays. Thus, differences in distribution of various markers within the tumor should be considered. Such differences were, in fact, encountered in the in situ hybridization analysis.

4.1. Galanin in corticotroph adenomas

A number of studies, using mainly immunostaining and in situ hybridization, have demonstrated presence of galanin in pituitary adenomas, especially in ACTH-producing tumors (see Table 3 for overview). Here we show that galanin levels in pituitary adenoma extracts vary with tumor type and correlate strongly with ACTH levels, supporting previous studies showing co-production of the two peptides. Indeed, the present in situ hybridization results using

Table 3

Galanin in human pituitary adenomas and in normal pituitaries

Reference	Pituitary adenomas				Normal pituitaries
	ACTH	GH	PRL	Inactive	
Hulting et al. [37]	—	1/1	—	—	—
Vrontakis et al. [38]	13/18	0/4	0/8	0/5	23/23
Hsu et al. [40]	16/19	5/11	2/14	9/18	7/7
Polak et al. [41]	6/6	3/8	2/2	0/7	—
Sano et al. [42]	9/14	0/10	0/9	0/9	—
Bennet et al. [39]	—	2/4	0/1	3/13	30/30
Invitti et al. [43]	5/5	—	—	—	—
Leung et al. [44]	10/16	1/26	1/19	25/89	6/6

Review of data published 1989-2002.

double-labeling on single sections show conclusively that individual cells can synthesize both ACTH and galanin. However, whereas coexistence of galanin and POMC mRNA was found in apparently all cells of one corticotroph adenoma (#1, 2, Table 1; Fig. 7A, B), another tumor of this type showed a clear separation of the two signals (#4, Table 1; Fig. 8A, C-E). Moreover, co-expression of galanin and POMC mRNA in the ACTH adenomas was reflected in a close correspondence between tumor content of galanin and ACTH, whereas when the peptides were expressed in separate cells RIAble tumor content of ACTH vastly exceeded that of galanin. Thus, absence of a correlation between these two peptides in extracts from corticotroph adenomas may reflect an expression of ACTH and galanin in different cell populations.

A clinical significance of this differential expression could not be related to age, sex, tumor size or pituitary function. However, as shown by Invitti et al. [43], the in vivo release of galanin in response to CRH varies among patients with Cushing's disease. We found no elevation of plasma galanin in sinus petrosus from two patients with Cushing's disease. Thus, although both patients showed a marked unilateral response to CRH, there was no similar effect on galanin levels. One of these patients (#4) was analyzed with double-labeling in situ hybridization, and in this patient galanin and ACTH were synthesized in distinctly separate cell populations (Fig. 8). Perhaps the variability in the response to CRH can be explained by the extent of colocalization of ACTH and galanin. Thus, when the peptides are produced in different cells, CRH may only cause release of ACTH. It would therefore be interesting to know whether or not type-1 CRH receptors [53] are present in ACTH-negative, galanin-positive cells. It should here be mentioned that in the normal state, stimulation of ACTH secretion with CRH, desmopressin or physical exercise does not affect plasma galanin levels [43,54,55].

The clinical significance of the presence of galanin in ACTH adenomas was also investigated by Leung et al. [44], showing that galanin-positive adenomas were smaller and had better prognosis than those not expressing galanin. The present study supports this finding; we have found that all seven corticotroph adenomas were galanin-positive, and

they were small and were successfully removed. Interestingly, in the four young patients with microadenomas and highly active Mb Cushing (#1–4) tumor galanin was inversely related to tumor volume. In contrast to Leung et al. [44] we found a correlation in six of seven corticotroph adenomas between galanin content and 24 h urinary cortisol excretion, representing integrated ACTH-cortisol production. To what extent galanin can serve as a marker for tumor activity as well as act as a growth inhibitory factor in this type of tumor requires analysis of a larger material.

4.2. Galanin in clinically inactive GH and PRL adenomas

The capacity of non-ACTH tumors to synthesize galanin has been controversial. Vrontakis et al. [38] and Sano et al. [42] found galanin-LI exclusively coexpressed with ACTH and not present in any other type of pituitary adenoma. We have previously demonstrated the presence of galanin in a GH adenoma [37], and Polak et al. [41] in GH and PRL adenomas. Hsu et al. [40], in addition, found galanin-LI in half of the non-functioning adenomas, including a few which were negative for ACTH. In the largest and most recent study, Leung et al. [44] demonstrated galanin-LI in one third of the clinically inactive adenomas but rarely in GH- or PRL-producing adenomas.

In the only previous study based on extraction, Bennet et al. [39], using an antiserum against porcine galanin and a porcine tracer, found detectable galanin in two of four GH adenomas and in three of thirteen clinically active adenomas. With the present extraction method we find galanin-LI in all but two of the non-ACTH adenomas studied. However, galanin levels above 5 pmol/mg protein were only seen in 5 of the 10 clinically inactive adenomas and in none of the GH- or PRL-producing adenomas.

The presence of galanin was shown in all normal pituitaries examined in previous studies [38–40,44]. In fact, extraction yielded markedly higher galanin concentrations in normal pituitaries than in non-ACTH pituitary adenomas [39]. Since the present study is based on tissue obtained at pituitary surgery, it does not include a control group of normal pituitaries. However, in a previous study of extracts from 280 postmortem pituitaries Schmidt et al. [5] reported 410 pmol galanin per gram frozen tissue. In our material the mean galanin content was 13,950 pmol/g frozen tissue in ACTH adenomas, 3.6 pmol in GH adenomas and 1.4–4170 pmol in clinically inactive adenomas.

The highly variable galanin content found in the clinically inactive adenomas correlated well with ACTH content in the same adenomas, and presence of normal corticotrophs in the samples containing high galanin and ACTH cannot be excluded. In one case (#19) this was verified by PAD showing normal pituitary tissue in 9 of 10 samples (cf. Fig. 9D–F). This patient also incurred a postoperative panhypopituitarism. Indeed, the three patients who became ACTH deficient after surgery (#18, 19, 24) had elevated tumor galanin-LI (mean 40.9 pmol/mg protein), whereas those with

preoperative ACTH failure (#10, 20, 21, 23) had low levels (mean 0.61 pmol/mg protein).

The patient (#16) with the highest tumor galanin content among the clinically inactive adenomas had shown clinical signs of Mb Cushing. However, laboratory investigation, performed after a tumor apoplexy, revealed normal ACTH and cortisol secretion. This patient had a recurrence of Cushing's disease a few years after surgery. Our sample may have contained ACTH adenomatous tissue or normal pituitary as indicated by immunocytochemistry.

4.3. Identity of galanin in pituitary tumors

The molecular form of galanin was studied with RP-HPLC in the seven corticotroph adenomas and the two clinically inactive adenomas with high galanin content (#1–7) (#16, 19). The elution pattern was shown to be remarkably homogenous with a small peak preceding the main form eluting with synthetic human galanin (Fig. 5A). It agrees with the pattern found in non-ACTH pituitary adenomas by Bennet et al. [39] and corresponds to the structure determined by Schmidt et al. [5] who isolated and sequenced galanin from purified extracts of postmortem pituitaries. Our finding indicates that the molecular form of galanin in corticotroph adenomas is identical to that found in clinically inactive tumors and in the normal pituitary. The small polar peak eluting earlier than galanin(1–30) probably represents degradation products which have been shown to elute prior to the main form [56]. In one case (#6) we found that galanin eluted later than human standard with a peak coinciding with galanin(1–16) (Fig. 5B). This fragment has not been found *in vivo* in the rat [56,57] or in man [2,5,28,58], and our finding suggests a different processing of tumor galanin in this patient.

4.4. Concluding remarks

We found that galanin is present in human corticotroph adenomas and provide conclusive evidence at the transcript level for cellular colocalization with POMC in some, but not all tumors. Our results also indicate that galanin present in other types of pituitary adenomas is associated with normal pituitary tissue. The molecular form of tumorous pituitary galanin is homogenous representing genuine human galanin (1–30), but in one adenoma a form of galanin coeluting with galanin(1–16) was identified. No evidence of galanin secretion from these adenomas into local or peripheral circulation was obtained, pointing to a mainly auto-and/or paracrine role for galanin produced in corticotroph adenomas.

Acknowledgements

The study was supported by grants from Stockholm County Council, Seidey and Rolf Fredriksson's Research

Foundation, the Swedish MRC (04X-2887), Marianne and Marcus Wallenberg's Foundation and an Unrestricted Bristol-Myers Squibb Neuroscience Grant. We are grateful to Agneta Hilding, Department of Molecular Medicine, Endocrine and Diabetes Unit, Karolinska Institutet for assistance with the statistical evaluation, and to Dr. Anders Höög for pathological anatomical diagnosis.

References

- [1] Tatemoto K, Rökaeus Å, Jönvall H, McDonald TJ, Mutt V. Galanin—a novel biologically active peptide from porcine intestine. *FEBS Lett* 1983;164:124–8.
- [2] Bersani M, Jølsen AH, Højrup P, Dunning BE, Andreasen JJ, Holst JJ. Human galanin: primary structure and identification of two molecular forms. *FEBS Lett* 1991;283:189–94.
- [3] Evans H, Shine J. Human galanin: molecular cloning reveals a unique structure. *Endocrinology* 1991;129:1682–4.
- [4] McKnight GL, Karlens AF, Kuwalyk S, Mathewes SL, Sheppard PO, O'Hara PJ, et al. Sequence of human galanin and its inhibition of glucose-stimulated insulin secretion from RIN cells. *Diabetes* 1991;41:82–7.
- [5] Schmidt WE, Kratzin II, Eckart K, Dreys D, Mundkowski G, Clemens A, et al. Isolation and primary structure of pituitary human galanin, a 30-residue nonamidated neuropeptide. *Proc Natl Acad Sci U S A* 1991;88:11435–9.
- [6] Vrontakis ME, Peden LM, Duckworth ML, Friesen HG. Isolation and characterization of a complementary DNA (galanin) clone from estrogen-induced pituitary tumor messenger RNA. *J Biol Chem* 1987;262:16755–8.
- [7] Kaplan LM, Gabriel SM, Koenig JL, Sundry ME, Spindel ER, Martin JB, et al. Galanin is an estrogen-inducible secretory product of the rat anterior pituitary. *Proc Natl Acad Sci U S A* 1988;85:7408–12.
- [8] Crawley JN. The role of galanin in feeding behaviour. *Neuropeptides* 1999;33:369–75.
- [9] Kalra SP, Horvath TL. Neuroendocrine interactions between galanin, opioids, and neuropeptide Y in the control of reproduction and appetite. *Ann N Y Acad Sci* 1998;863:236–40.
- [10] Leibowitz S. Differential functions of hypothalamic galanin cell groups in the regulation of eating and body weight. *Ann N Y Acad Sci* 1998;863:206–20.
- [11] McDonald MP, Gleason TC, Robinson JK, Crawley J. Galanin inhibits performance on rodent memory tasks. In: Hökfelt T, Bartfai T, Crawley J, editors. *Galanin basic research discoveries and therapeutic implications*. New York: NY Acad. Sci.; 1998. p. 206–20.
- [12] Ögren S-O, Schöft PA, Kehr J, Yoshitake T, Misane I, Mannström P, et al. Modulation of acetylcholine and serotonin transmission by galanin. *Ann N Y Acad Sci U S A* 1998;863:342–63.
- [13] Xu XJ, Hökfelt T, Bartfai T, Wiesenfeld-Hallin Z. Galanin and spinal nociceptive mechanisms: recent advances and therapeutic implications. *Neuropeptides* 2000;34:137–47.
- [14] Habert-Ortoli E, Amiranoff B, Loquet I, Laburthe M, Mayaux JF. Molecular cloning of a functional human galanin receptor. *Proc Natl Acad Sci U S A* 1994;91:9780–3.
- [15] Branchek TA, Smith KE, Gerald C, Walker MW. Galanin receptor subtypes. *Trends Pharmacol Sci* 2000;21:109–16.
- [16] Lisma TP, Shine J. Galanin and galanin receptors. *Results Probl Cell Differ* 1999;26:257–91.
- [17] Wynick D, Small CJ, Bacon A, Holmes FE, Norman M, Ormandy CJ, et al. Galanin regulates prolactin release and lactotroph proliferation. *Proc Natl Acad Sci U S A* 1998;95:12671–6.
- [18] Cai A, Hayes JD, Patel N, Hyde JF. Targeted overexpression of galanin in lactotrophs of transgenic mice induces hyperprolactinemia and pituitary hyperplasia. *Endocrinology* 1999;140:4955–64.
- [19] Kokaia M, Holmberg K, Nanobashvili A, Xu Z-QD, Kokaia Z, Lendahl U, et al. Suppressed kindling epileptogenesis in mice with ectopic overexpression of galanin. *Proc Natl Acad Sci U S A* 2001;98:14006–11.
- [20] Steiner RA, Hohmann JG, Holmes A, Wrenn CC, Cadd G, Jureus A, et al. Galanin transgenic mice display cognitive and neurochemical deficits characteristic of Alzheimer's disease. *Proc Natl Acad Sci U S A* 2001;98:4184–9.
- [21] Rökaeus Å, Melander T, Hökfelt T, Lundberg JM, Tatemoto K, Carlquist M, et al. A galanin-like peptide in the central nervous system and intestine of the rat. *Neurosci Lett* 1984;47:161–6.
- [22] Skofitsch G, Jacobowitz DM. Immunohistochemical mapping of galanin-like neurons in the rat central nervous system. *Peptides* 1985;6:509–46.
- [23] Skofitsch G, Jacobowitz DM. Quantitative distribution of galanin-like immunoreactivity in the rat central nervous system. *Peptides* 1986;7:609–13.
- [24] Melander T, Hökfelt T, Rökaeus Å. Distribution of galanin-like immunoreactivity in the rat central nervous system. *J Comp Neurol* 1986;248:475–517.
- [25] Bloch GJ, Butler PC, Eckersell CB, Mills RH. Gonadal steroid-dependent GAL-IR cells within the medial preoptic nucleus (MPN) and the stimulatory effects of GAL within the MPN on sexual behaviors. *Ann N Y Acad Sci U S A* 1998;863:188–205.
- [26] Hohmann JG, Clifton DK, Steiner RA. Galanin: Analysis of its coexpression in gonadotrophin-releasing hormone and growth hormone-releasing hormone neurons. *Ann N Y Acad Sci U S A* 1998;863:221–35.
- [27] Merchenthaler I. LHRH and sexual dimorphism. *Ann N Y Acad Sci U S A* 1998;863:175–87.
- [28] Bauer FF, Ginsberg I, Venetikonu M, MacKay DJ, Burin JM, Bloom SR. Growth hormone release in man induced by galanin, a new hypothalamic peptide. *Lancet* 1986;26:192–5.
- [29] Davis TME, Burrin JM, Bloom SR. Growth hormone release in response to GHRH-releasing hormone in man is 3-fold enhanced by galanin. *J Clin Endocrinol Metab* 1987;65:1248–52.
- [30] Cimino V. Galanin inhibits ACTH release in vitro and can be demonstrated immunocytochemically in dispersed corticotrophs. *Exp Cell Res* 1996;228:212–5.
- [31] Giustina A, Licini M, Schettino M, Doga M, Pizzoccolo G, Negro-Villar A. Physiological role of galanin in the regulation of anterior pituitary function in humans. *Am J Physiol* 1994;266:E57–61.
- [32] Hooi SC, Maiter DM, Martin JB, Koenig JL. Galaninergic mechanisms are involved in the regulation of corticotropin and thyrotropin secretion in the rat. *Endocrinology* 1990;127:2281–9.
- [33] Hooi SC, Koenig JL, Gabriel SM, Maiter D, Martin JB. Influence of thyroid hormone on the concentration of galanin in the rat brain and pituitary. *Neuroendocrinology* 1990;51:351–6.
- [34] Wynick D, Smith DM, Ghatei M, Akinsanya K, Bhogal R, Purkiss P, et al. Characterization of a high-affinity galanin receptor in the rat anterior pituitary: Absence of biological effect and reduced membrane binding of the antagonist M15 differentiate it from the brain/gut receptor. *Proc Natl Acad Sci U S A* 1993;90:4231–5.
- [35] Wynick D, Hammond PJ, Akinsanya KO, Bloom SR. Galanin regulates basal and oestrogen-stimulated lactotroph function. *Nature* 1993;364:529–32.
- [36] Moore JPI, Morrison DG, Hyde JF. Galanin gene expression is increased in the anterior pituitary gland of the human growth hormone-releasing hormone transgenic mice. *Endocrinology* 1994;134:2005–10.
- [37] Hulting A-L, Meister B, Grimelius L, Wersäll J, Ånggård A, Hökfelt T. Production of a galanin-like peptide by a human pituitary adenoma: immunohistochemical evidence. *Acta Physiol Scand* 1989;137:561–2.
- [38] Vrontakis ME, Sano T, Kovacs K, Friesen HG. Presence of galanin-like immunoreactivity in non-tumorous corticotrophs and corticotroph adenomas of the human pituitary. *J Clin Endocrinol Metab* 1990;70:747–51.

- [39] Bennet WM, Hill SF, Ghatei MA, Bloom SR. Galanin in the normal pituitary and brain and in pituitary adenomas. *J Endocrinol* 1991; 130:463-7.
- [40] Hsu DW, Hoon SC, Hedley-Whyte T, Strauss RM, Kaplan LM. Coexpression of galanin and adrenocorticotrophic hormone in human pituitary and pituitary adenomas. *Am J Pathol* 1991;138:897-909.
- [41] Polak JM, Gibson S, Gentleman S, Steel J, Van Noorden S. Galanin: distribution, ontogeny and expression following manipulation of the endocrine and nervous system. In: Hökfelt T, Bartfai T, Jacobowitz D., Ottoson D., editors. *Galanin a new multifunctional peptide in the neuro-endocrine system*. London: Macmillan; 1991. p. 117-34.
- [42] Sano T, Vrontakis MF, Kovacs K, Asa SL, Friesen HG. Galanin immunoreactivity in neuroendocrine tumors. *Arch Pathol Lab Med* 1991;115:926-9.
- [43] Invitti C, Pecori F, Girakli P, Dubini A, Muroi P, Losa M, et al. Galanin is released by adrenocorticotropin-secreting pituitary adenomas in vivo and in vitro. *J Clin Endocrinol Metab* 1999;4:1351-6.
- [44] Leung B, Iismaa TP, Leung K-C, Ilori YJ, Turner J, Sheehy JP, et al. Galanin in human pituitary adenomas: frequency and clinical significance. *Clin Endocrinol* 2002;56:397-403.
- [45] Grenbäck E, Bjellerup P, Hulting A-I, Lundblad L, Änggård A. Galanin in hypophysadenoma. *Proc Ann Meeting Swedish Med Soc* 1997;29P:162.
- [46] P. Lundin, *Magnetic resonance imaging of pituitary macroadenomas*. Thesis, Uppsala ISBN 91-628-0511-8 1992.
- [47] Wilson CB. A decade of pituitary microsurgery. The Herbert Olivecrona lecture. *J Neurosurg* 1984;61:314-33.
- [48] Schmidt WE, Siegel EG, Lamberts R, Gallwitz B, Creutzfeldt W. Pancreastatin: molecular and immunocytochemical characterization of a novel peptide in porcine and human tissues. *Endocrinology* 1988;123:1395-404.
- [49] Drouin J, Goodman HM. Most of the coding region of rat ACTH beta-LPH precursor gene lacks intervening sequences. *Nature* 1980; 288:610-3.
- [50] Dagerlind Å, Friberg K, Bean A, Hökfelt T. Sensitive mRNA detection using unfixed tissue: combined radioactive and non-radioactive in situ hybridization histochemistry. *Histochemistry* 1992; 98:39-49.
- [51] Schnitz GG, Walter T, Seibl R, Kessler C. Nonradioactive labeling of oligonucleotides in vitro with the hapten digoxigenin by tailing with terminal transferase. *Ann Biochem* 1991;192:222-31.
- [52] Broberger C, Landry M, Wong H, Walsh J, Hökfelt T. Subtypes Y1 and Y2 of the neuropeptide Y receptor are respectively expressed in proopiomelanocortin and neuropeptide Y-containing neurons of the rat hypothalamic arcuate nucleus. *Neuroendocrinology* 1997;66: 393-408.
- [53] Aguilera G, Rabadan-Diehl C, Nikodemova M. Regulation of pituitary corticotropin releasing hormone receptors. *Peptides* 2001;22: 769-74.
- [54] Ceresini G, Freddi M, Paccotti P, Valenti G, Merchenthaler I. Effects of ovine corticotropin-releasing hormone injection and desmopressin coadministration on galanin and adrenotropin plasma levels in normal women. *J Clin Endocrinol Metab* 1997;82:607-10.
- [55] Ceresini G, Marchini L, Fabbro A, Freddi M, Pasolini G, Reali N, et al. Evaluation of circulating galanin levels after exercise-induced pituitary hormone secretion in man. *Metabolism* 1997;82:282-6.
- [56] Land T, Langel U, Bartfai T. Hypothalamic degradation of galanin (1-29) and galanin (1-16): identification and characterization of the peptidolytic products. *Brain Res* 1991;558:245-50.
- [57] Hedecs K, Langel Ü, Xu X-J, Wiesenfeld-Hallin Z, Bartfai T. Biological activities of two endogenously occurring N-terminally extended forms of galanin in the rat spinal cord. *Eur J Pharmacol* 1994;259: 151-6.
- [58] Bauer FE, Hacker GW, Terenghi G, Adrian TE, Polak JM, Bloom SR. Localization and molecular forms of galanin in human adrenals: elevated levels in pheochromocytomas. *J Clin Endocrinol Metab* 1986; 63:1372.

GENES

Benjamin Lewin

Oxford New York Tokyo
Oxford University Press
1997

*Oxford University Press, Great Clarendon Street, Oxford OX2 6DP
Oxford New York
Athens Auckland Bangkok Bogota Bombay Buenos Aires
Calcutta Cape Town Dar es Salaam Delhi Florence Hong Kong
Istanbul Karachi Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore Taipei Tokyo Toronto
and associated companies in
Berlin Ibadan*

Oxford is a trade mark of Oxford University Press

*Published in the United States
by Oxford University Press, Inc., New York*

© Oxford University Press and Cell Press, 1997

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press. Within the U.K. exceptions are allowed in respect of any fair dealing for the purpose of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms and in other countries should be sent to the Rights Department, Oxford University Press, at the address above.

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging in Publication Data
(Data available)*

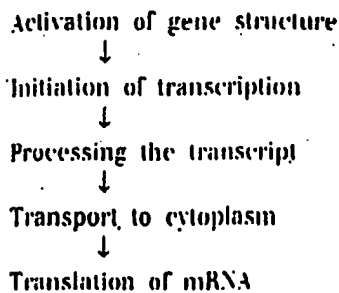
Typeset by Hyvern Typesetting Ltd, Bristol

Printed in The United States of America

CHAPTER 29

Regulation of transcription

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:



The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is

controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing: some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein; it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide

common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor*. An element that causes a gene to respond to such a factor is called a response element; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of

the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCGNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Review

Translation Initiation in Cancer: A Novel Target for Therapy¹

Funda Meric² and Kelly K. Hunt

Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract

Translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled with cell cycle progression and cell growth. Several alterations in translational control occur in cancer. Variant mRNA sequences can alter the translational efficiency of individual mRNA molecules, which in turn play a role in cancer biology. Changes in the expression or availability of components of the translational machinery and in the activation of translation through signal transduction pathways can lead to more global changes, such as an increase in the overall rate of protein synthesis and translational activation of the mRNA molecules involved in cell growth and proliferation. We review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to help elucidate new therapeutic avenues.

Introduction

The fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells. With the advent of cDNA array technology, most efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable either to DNA amplification or to differences in transcription. Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.

The power of translational regulation has been best recognized among developmental biologists, because transcription does not occur in early embryogenesis in eukaryotes. For example, in *Xenopus*, the period of transcriptional quiescence continues until the embryo reaches midblastula transition, the 4000-cell stage. Therefore, all necessary mRNA molecules are transcribed during oogenesis and stockpiled in a translationally inactive, masked form. The mRNA are translationally activated at appropriate times during oocyte maturation, fertilization, and

early embryogenesis and thus, are under strict translational control.

Translation has an established role in cell growth. Basically, an increase in protein synthesis occurs as a consequence of mitogenesis. Until recently, however, little was known about the alterations in mRNA translation in cancer, and much is yet to be discovered about their role in the development and progression of cancer. Here we review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to elucidate potential new therapeutic avenues.

Basic Principles of Translational Control

Mechanism of Translation Initiation

Translation initiation is the main step in translational regulation. Translation initiation is a complex process in which the initiator tRNA and the 40S and 60S ribosomal subunits are recruited to the 5' end of a mRNA molecule and assembled by eukaryotic translation initiation factors into an 80S ribosome at the start codon of the mRNA (Fig. 1). The 5' end of eukaryotic mRNA is capped, i.e., contains the cap structure m⁷GpppN (7-methyl-guanosine-triphospho-5'-ribonucleoside). Most translation in eukaryotes occurs in a cap-dependent fashion, i.e., the cap is specifically recognized by the eIF4E,³ which binds the 5' cap. The eIF4F translation initiation complex is then formed by the assembly of eIF4E, the RNA helicase eIF4A, and eIF4G, a scaffolding protein that mediates the binding of the 40S ribosomal subunit to the mRNA molecule through interaction with the eIF3 protein present on the 40S ribosome. eIF4A and eIF4B participate in melting the secondary structure of the 5' UTR of the mRNA. The 43S initiation complex (40S/eIF2/Met-tRNA/GTP complex) scans the mRNA in a 5'→3' direction until it encounters an AUG start codon. This start codon is then base-paired to the anticodon of initiator tRNA, forming the 48S initiation complex. The initiation factors are then displaced from the 48S complex, and the 60S ribosome joins to form the 80S ribosome.

Unlike most eukaryotic translation, translation initiation of certain mRNAs, such as the picornavirus RNA, is cap independent and occurs by internal ribosome entry. This mechanism does not require eIF4E. Either the 43S complex can bind the initiation codon directly through interaction with the IRES in the 5' UTR such as in the encephalomyocarditis virus, or it can

Received 5/16/02; revised 7/12/02; accepted 7/22/02.

¹ F. M. is supported by The University of Texas M. D. Anderson Cancer Center Physician-Scientist Program and by NIH Grant 1K08-CA 91895-01. K. K. H. is supported by Department of Defense Award DAMD-17-97-1-7162.

² To whom requests for reprints should be addressed, at Department of Surgical Oncology, Box 444, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 745-4453; Fax: (713) 745-4926; E-mail: fmeric@mdanderson.org.

³ The abbreviations used are: eIF4E, eukaryotic initiation factor 4E; UTR, untranslated region; IRES, internal ribosome entry site; 4E-BP1, eukaryotic initiation factor 4E-binding protein 1; S6K, ribosomal p70 S6 kinase; mTOR, mammalian target of rapamycin; ATM, ataxia telangiectasia mutated; PI3K, phosphatidylinositol 3-kinase; PTEN, phosphatase and tensin homolog deleted from chromosome 10; PP2A, protein phosphatase 2A; TGF- β , transforming growth factor- β ; PAP, poly(A) polymerase; EPA, eicosapentaenoic acid; mda-7, melanoma differentiation-associated gene 7.

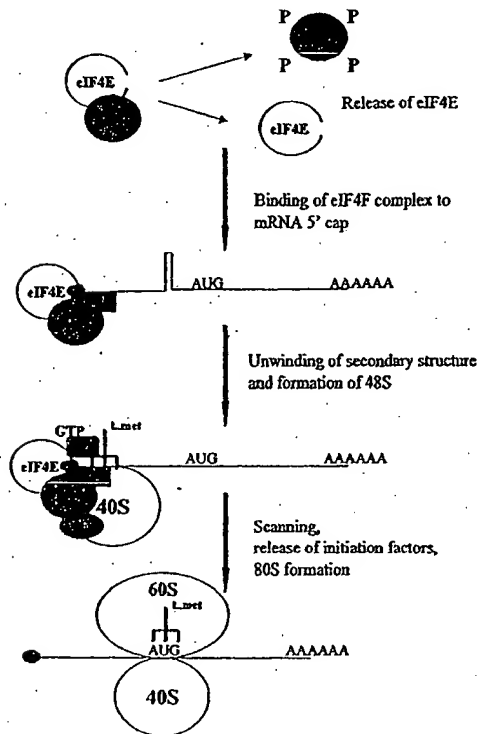


Fig. 1. Translation initiation in eukaryotes. The 4E-BPs are hyperphosphorylated to release eIF4E so that it can interact with the 5' cap, and the eIF4F initiation complex is assembled. The interaction of poly(A) binding protein with the initiation complex and circularization of the mRNA is not depicted in the diagram. The secondary structure of the 5' UTR is melted, the 40S ribosomal subunit is bound to eIF3, and the ternary complex consisting of eIF2, GTP, and the Met-tRNA are recruited to the mRNA. The ribosome scans the mRNA in a 5'→3' direction until an AUG start codon is found in the appropriate sequence context. The initiation factors are released, and the large ribosomal subunit is recruited.

initially attach to the IRES and then reach the initiation codon by scanning or transfer, as is the case with the poliovirus (1).

Regulation of Translation Initiation

Translation initiation can be regulated by alterations in the expression or phosphorylation status of the various factors involved. Key components in translational regulation that may provide potential therapeutic targets follow.

eIF4E. eIF4E plays a central role in translation regulation. It is the least abundant of the initiation factors and is considered the rate-limiting component for initiation of cap-dependent translation. eIF4E may also be involved in mRNA splicing, mRNA 3' processing, and mRNA nucleocytoplasmic transport (2). eIF4E expression can be increased at the transcriptional level in response to serum or growth factors (3). eIF4E overexpression may cause preferential translation of mRNAs containing excessive secondary structure in their 5' UTR that are normally discriminated against by the trans-

lational machinery and thus are inefficiently translated (4–7). As examples of this, overexpression of eIF4E promotes increased translation of vascular endothelial growth factor, fibroblast growth factor-2, and cyclin D1 (2, 8, 9).

Another mechanism of control is the regulation of eIF4E phosphorylation. eIF4E phosphorylation is mediated by the mitogen-activated protein kinase-interacting kinase 1, which is activated by the mitogen-activated pathway, activating extracellular signal-related kinases and the stress-activated pathway acting through p38 mitogen-activated protein kinase (10–13). Several mitogens, such as serum, platelet-derived growth factor, epidermal growth factor, insulin, angiotensin II, src kinase overexpression, and ras overexpression, lead to eIF4E phosphorylation (14). The phosphorylation status of eIF4E is usually correlated with the translational rate and growth status of the cell; however, eIF4E phosphorylation has also been observed in response to some cellular stresses when translational rates actually decrease (15). Thus, further study is needed to understand the effects of eIF4E phosphorylation on eIF4E activity.

Another mechanism of regulation is the alteration of eIF4E availability by the binding of eIF4E to the eIF4E-binding proteins (4E-BP, also known as PHAS-I). 4E-BPs compete with eIF4G for a binding site in eIF4E. The binding of eIF4E to the best characterized eIF4E-binding protein, 4E-BP1, is regulated by 4E-BP1 phosphorylation. Hypophosphorylated 4E-BP1 binds to eIF4E, whereas 4E-BP1 hyperphosphorylation decreases this binding. Insulin, angiotensin, epidermal growth factor, platelet-derived growth factor, hepatocyte growth factor, nerve growth factor, insulin-like growth factors I and II, interleukin 3, granulocyte-macrophage colony-stimulating factor + steel factor, gastrin, and the adenovirus have all been reported to induce phosphorylation of 4E-BP1 and to decrease the ability of 4E-BP1 to bind eIF4E (15, 16). Conversely, deprivation of nutrients or growth factors results in 4E-BP1 dephosphorylation, an increase in eIF4E binding, and a decrease in cap-dependent translation.

p70 S6 Kinase. Phosphorylation of ribosomal 40S protein S6 by S6K is thought to play an important role in translational regulation. S6K $-/-$ mouse embryonic cells proliferate more slowly than do parental cells, demonstrating that S6K has a positive influence on cell proliferation (17). S6K regulates the translation of a group of mRNAs possessing a 5' terminal oligopyrimidine tract (5' TOP) found at the 5' UTR of ribosomal protein mRNAs and other mRNAs coding for components of the translational machinery. Phosphorylation of S6K is regulated in part based on the availability of nutrients (18, 19) and is stimulated by several growth factors, such as platelet-derived growth factor and insulin-like growth factor I (20).

eIF2 α Phosphorylation. The binding of the initiator tRNA to the small ribosomal unit is mediated by translation initiation factor eIF2. Phosphorylation of the α -subunit of eIF2 prevents formation of the eIF2/GTP/Met-tRNA complex and inhibits global protein synthesis (21, 22). eIF2 α is phosphorylated under a variety of conditions, such as viral infection, nutrient deprivation, heme deprivation, and apoptosis (22). eIF2 α is phosphorylated by heme-regulated inhibitor, nutrient-regulated protein kinase, and the IFN-induced, double-stranded RNA-activated protein kinase (PKR; Ref. 23).

The mTOR Signaling Pathway. The macrolide antibiotic rapamycin (Siralimus; Wyeth-Ayerst Research, Collegeville, PA) has been the subject of intensive study because it inhibits signal transduction pathways involved in T-cell activation. The rapamycin-sensitive component of these pathways is mTOR (also called FRAP or RAFT1). mTOR is the mammalian homologue of the yeast TOR proteins that regulate G₁ progression and translation in response to nutrient availability (24). mTOR is a serine-threonine kinase that modulates translation initiation by altering the phosphorylation status of 4E-BP1 and S6K (Fig. 2; Ref. 25).

4E-BP1 is phosphorylated on multiple residues. mTOR phosphorylates the Thr-37 and Thr-46 residues of 4E-BP1 *in vitro* (26); however, phosphorylation at these sites is not associated with a loss of eIF4E binding. Phosphorylation of Thr-37 and Thr-46 is required for subsequent phosphorylation at several COOH-terminal, serum-sensitive sites; a combination of these phosphorylation events appears to be needed to inhibit the binding of 4E-BP1 to eIF4E (25). The product of the ATM gene, p38/MSK1 pathway, and protein kinase C α also play a role in 4E-BP1 phosphorylation (27–29).

S6K and 4E-BP1 are also regulated, in part, by PI3K and its downstream protein kinase Akt. PTEN is a phosphatase that negatively regulates PI3K signaling. PTEN null cells have constitutively active Akt, with increased S6K activity and S6 phosphorylation (30). S6K activity is inhibited both by PI3K inhibitors wortmannin and LY294002 and by mTOR inhibitor rapamycin (24). Akt phosphorylates Ser-2448 in mTOR *in vitro*, and this site is phosphorylated upon Akt activation *in vivo* (31–33). Thus, mTOR is regulated by the PI3K/Akt pathway; however, this does not appear to be the only mode of regulation of mTOR activity. Whether the PI3K pathway also regulates S6K and 4E-BP1 phosphorylation independent of mTOR is controversial.

Interestingly, mTOR autophosphorylation is blocked by wortmannin but not by rapamycin (34). This seeming inconsistency suggests that mTOR-responsive regulation of 4E-BP1 and S6K activity occurs through a mechanism other than intrinsic mTOR kinase activity. An alternate pathway for 4E-BP1 and S6K phosphorylation by mTOR activity is by the inhibition of a phosphatase. Treatment with calyculin A, an inhibitor of phosphatases 1 and 2A, reduces rapamycin-induced dephosphorylation of 4E-BP1 and S6K by rapamycin (35). PP2A interacts with full-length S6K but not with a S6K mutant that is resistant to dephosphorylation resulting from rapamycin. mTOR phosphorylates PP2A *in vitro*; however, how this process alters PP2A activity is not known. These results are consistent with the model that phosphorylation of a phosphatase by mTOR prevents dephosphorylation of 4E-BP1 and S6K, and conversely, that nutrient deprivation and rapamycin block inhibition of the phosphatase by mTOR.

Polyadenylation. The poly(A) tail in eukaryotic mRNA is important in enhancing translation initiation and mRNA stability. Polyadenylation plays a key role in regulating gene expression during oogenesis and early embryogenesis. Some mRNA that are translationally inactive in the oocyte are polyadenylated concomitantly with translational activation in oocyte maturation, whereas other mRNAs that are translationally active during oogenesis are deadenylated and trans-

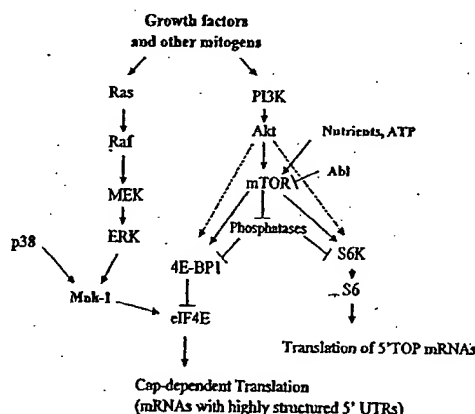


Fig. 2. Regulation of translation initiation by signal transduction pathways. Signaling via p38, extracellular signal-related kinase, PI3K, and mTOR can all activate translation initiation.

lationally silenced (36–38). Thus, control of poly(A) tail synthesis is an important regulatory step in gene expression. The 5' cap and poly(A) tail are thought to function synergistically to regulate mRNA translational efficiency (39, 40).

RNA Packaging. Most RNA-binding proteins are assembled on a transcript at the time of transcription, thus determining the translational fate of the transcript (41). A highly conserved family of Y-box proteins is found in cytoplasmic messenger ribonucleoprotein particles, where the proteins are thought to play a role in restricting the recruitment of mRNA to the translational machinery (41–43). The major mRNA-associated protein, YB-1, destabilizes the interaction of eIF4E and the 5' mRNA cap *in vitro*, and overexpression of YB-1 results in translational repression *in vivo* (44). Thus, alterations in RNA packaging can also play an important role in translational regulation.

Translation Alterations Encountered in Cancer

Three main alterations at the translational level occur in cancer: variations in mRNA sequences that increase or decrease translational efficiency, changes in the expression or availability of components of the translational machinery, and activation of translation through aberrantly activated signal transduction pathways. The first alteration affects the translation of an individual mRNA that may play a role in carcinogenesis. The second and third alterations can lead to more global changes, such as an increase in the overall rate of protein synthesis, and the translational activation of several mRNA species.

Variations in mRNA Sequence

Variations in mRNA sequence affect the translational efficiency of the transcript. A brief description of these variations and examples of each mechanism follow.

Mutations. Mutations in the mRNA sequence, especially in the 5' UTR, can alter its translational efficiency, as seen in the following examples.

c-myc. Saito *et al.* proposed that translation of full-length *c-myc* is repressed, whereas in several Burkitt lymphomas that have deletions of the mRNA 5' UTR, translation of *c-myc* is more efficient (45). More recently, it was reported that the 5' UTR of *c-myc* contains an IRES, and thus *c-myc* translation can be initiated by a cap-independent as well as a cap-dependent mechanism (46, 47). In patients with multiple myeloma, a C→T mutation in the *c-myc* IRES was identified (48) and found to cause an enhanced initiation of translation via internal ribosomal entry (49).

BRCA1. A somatic point mutation (117 G→C) in position -3 with respect to the start codon of the *BRCA1* gene was identified in a highly aggressive sporadic breast cancer (50). Chimeric constructs consisting of the wild-type or mutated *BRCA1* 5' UTR and a downstream luciferase reporter demonstrated a decrease in the translational efficiency with the 5' UTR mutation.

Cyclin-dependent Kinase Inhibitor 2A. Some inherited melanoma kindreds have a G→T transversion at base -34 of cyclin-dependent kinase inhibitor-2A, which encodes a cyclin-dependent kinase 4/cyclin-dependent kinase 6 kinase inhibitor important in G₁ checkpoint regulation (51). This mutation gives rise to a novel AUG translation initiation codon, creating an upstream open reading frame that competes for scanning ribosomes and decreases translation from the wild-type AUG.

Alternate Splicing and Alternate Transcription Start Sites. Alterations in splicing and alternate transcription sites can lead to variations in 5' UTR sequence, length, and secondary structure, ultimately impacting translational efficiency.

ATM. The *ATM* gene has four noncoding exons in its 5' UTR that undergo extensive alternative splicing (52). The contents of 12 different 5' UTRs that show considerable diversity in length and sequence have been identified. These divergent 5' leader sequences play an important role in the translational regulation of the *ATM* gene.

mdm. In a subset of tumors, overexpression of the oncoprotein *mdm2* results in enhanced translation of the *mdm2* mRNA. Use of different promoters leads to two *mdm2* transcripts that differ only in their 5' leaders (53). The longer 5' UTR contains two upstream open reading frames, and this mRNA is loaded with ribosomes inefficiently compared with the short 5' UTR.

BRCA1. In a normal mammary gland, *BRCA1* mRNA is expressed with a shorter leader sequence (5' UTRa), whereas in sporadic breast cancer tissue, *BRCA1* mRNA is expressed with a longer leader sequence (5' UTRb); the translational efficiency of transcripts containing 5' UTRb is 10 times lower than that of transcripts containing 5' UTRa (54).

TGF-β3. *TGF-β3* mRNA includes a 1.1-kb 5' UTR, which exerts an inhibitory effect on translation. Many human breast cancer cell lines contain a novel *TGF-β3* transcript with a 5' UTR that is 870 nucleotides shorter and has a 7-fold greater translational efficiency than the normal *TGF-β3* mRNA (55).

Alternate Polyadenylation Sites. Multiple polyadenylation signals leading to the generation of several transcripts with differing 3' UTR have been described for several mRNA species, such as the *RET* proto-oncogene (56), *ATM* gene (52), tissue inhibitor of metalloproteinases-3 (57), *RHOA*

proto-oncogene (58), and calmodulin-1 (59). Although the effect of these alternate 3' UTRs on translation is not yet known, they may be important in RNA-protein interactions that affect translational recruitment. The role of these alterations in cancer development and progression is unknown.

Alterations in the Components of the Translation Machinery

Alterations in the components of translation machinery can take many forms.

Overexpression of eIF4E. Overexpression of eIF4E causes malignant transformation in rodent cells (60) and the deregulation of HeLa cell growth (61). Polunovsky *et al.* (62) found that eIF4E overexpression substitutes for serum and individual growth factors in preserving viability of fibroblasts, which suggests that eIF4E can mediate both proliferative and survival signaling.

Elevated levels of eIF4E mRNA have been found in a broad spectrum of transformed cell lines (63). eIF4E levels are elevated in all ductal carcinoma *in situ* specimens and invasive ductal carcinomas, compared with benign breast specimens evaluated with Western blot analysis (64, 65). Preliminary studies suggest that this overexpression is attributable to gene amplification (66).

There are accumulating data suggesting that eIF4E overexpression can be valuable as a prognostic marker. eIF4E overexpression was found in a retrospective study to be a marker of poor prognosis in stages I to III breast carcinoma (67). Verification of the prognostic value of eIF4E in breast cancer is now under way in a prospective trial (67). However, in a different study, eIF4E expression was correlated with the aggressive behavior of non-Hodgkin's lymphomas (68). In a prospective analysis of patients with head and neck cancer, elevated levels of eIF4E in histologically tumor-free surgical margins predicted a significantly increased risk of local-regional recurrence (9). These results all suggest that eIF4E overexpression can be used to select patients who might benefit from more aggressive systemic therapy. Furthermore, the head and neck cancer data suggest that eIF4E overexpression is a field defect and can be used to guide local therapy.

Alterations in Other Initiation Factors. Alterations in a number of other initiation factors have been associated with cancer. Overproduction of eIF4G, similar to eIF4E, leads to malignant transformation *in vitro* (69). eIF-2α is found in increased levels in bronchioloalveolar carcinomas of the lung (3). Initiation factor eIF-4A1 is overexpressed in melanoma (70) and hepatocellular carcinoma (71). The p40 subunit of translation initiation factor 3 is amplified and overexpressed in breast and prostate cancer (72), and the eIF3-p110 subunit is overexpressed in testicular seminoma (73). The role that overexpression of these initiation factors plays on the development and progression of cancer, if any, is not known.

Overexpression of S6K. S6K is amplified and highly overexpressed in the MCF7 breast cancer cell line, compared with normal mammary epithelium (74). In a study by Bartlund *et al.* (74), S6K was amplified in 59 of 668 primary breast tumors, and a statistically significant association was observed between amplification and poor prognosis.

Overexpression of PAP. PAP catalyzes 3' poly(A) synthesis. PAP is overexpressed in human cancer cells compared with normal and virally transformed cells (75). PAP enzymatic activity in breast tumors has been correlated with PAP protein levels (76) and, in mammary tumor cytosols, was found to be an independent factor for predicting survival (76). Little is known, however, about how PAP expression or activity affects the translational profile.

Alterations in RNA-binding Proteins. Even less is known about alterations in RNA packaging in cancer. Increased expression and nuclear localization of the RNA-binding protein YB-1 are indicators of a poor prognosis for breast cancer (77), non-small cell lung cancer (78), and ovarian cancer (79). However, this effect may be mediated at least in part at the level of transcription, because YB-1 increases chemoresistance by enhancing the transcription of a multidrug resistance gene (80).

Activation of Signal Transduction Pathways

Activation of signal transduction pathways by loss of tumor suppressor genes or overexpression of certain tyrosine kinases can contribute to the growth and aggressiveness of tumors. An important mutant in human cancers is the tumor suppressor gene *PTEN*, which leads to the activation of the PI3K/Akt pathway. Activation of PI3K and Akt induces the oncogenic transformation of chicken embryo fibroblasts. The transformed cells show constitutive phosphorylation of S6K and of 4E-BP1 (81). A mutant Akt that retains kinase activity but does not phosphorylate S6K or 4E-BP1 does not transform fibroblasts, which suggests a correlation between the oncogenicity of PI3K and Akt and the phosphorylation of S6K and 4E-BP1 (81).

Several tyrosine kinases such as platelet-derived growth factor, insulin-like growth factor, HER2/neu, and epidermal growth factor receptor are overexpressed in cancer. Because these kinases activate downstream signal transduction pathways known to alter translation initiation, activation of translation is likely to contribute to the growth and aggressiveness of these tumors. Furthermore, the mRNA for many of these kinases themselves are under translational control. For example, HER2/neu mRNA is translationally controlled both by a short upstream open reading frame that represses HER2/neu translation in a cell type-independent manner and by a distinct cell type-dependent mechanism that increases translational efficiency (82). HER2/neu translation is different in transformed and normal cells. Thus, it is possible that alterations at the translational level can, in part, account for the discrepancy between *HER2/neu* gene amplification detected by fluorescence *in situ* hybridization and protein levels detected by immunohistochemical assays.

Translation Targets of Selected Cancer Therapy

Components of the translation machinery and signal pathways involved in the activation of translation initiation represent good targets for cancer therapy.

Targeting the mTOR Signaling Pathway: Rapamycin and Temstatin

Rapamycin inhibits the proliferation of lymphocytes. It was initially developed as an immunosuppressive drug for organ

transplantation. Rapamycin with FKBP 12 (FK506-binding protein, *M*, 12,000) binds to mTOR to inhibit its function.

Rapamycin causes a small but significant reduction in the initiation rate of protein synthesis (83). It blocks cell growth in part by blocking S6 phosphorylation and selectively suppressing the translation of 5' TOP mRNAs, such as ribosomal proteins, and elongation factors (83–85). Rapamycin also blocks 4E-BP1 phosphorylation and inhibits cap-dependent but not cap-independent translation (17, 86).

The rapamycin-sensitive signal transduction pathway, activated during malignant transformation and cancer progression, is now being studied as a target for cancer therapy (87). Prostate, breast, small cell lung, glioblastoma, melanoma, and T-cell leukemia are among the cancer lines most sensitive to the rapamycin analogue CCI-779 (Wyeth-Ayerst Research; Ref. 87). In rhabdomyosarcoma cell lines, rapamycin is either cytostatic or cytotoxic, depending on the p53 status of the cell; p53 wild-type cells treated with rapamycin arrest in the G₁ phase and maintain their viability, whereas p53 mutant cells accumulate in G₁ and undergo apoptosis (88, 89). In a recently reported study using human primitive neuroectodermal tumor and medulloblastoma models, rapamycin exhibited more cytotoxicity in combination with cisplatin and camptothecin than as a single agent. *In vivo*, CCI-779 delayed growth of xenografts by 160% after 1 week of therapy and 240% after 2 weeks. A single high-dose administration caused a 37% decrease in tumor volume. Growth inhibition *in vivo* was 1.3 times greater, with cisplatin in combination with CCI-779 than with cisplatin alone (90). Thus, preclinical studies suggest that rapamycin analogues are useful as single agents and in combination with chemotherapy.

Rapamycin analogues CCI-779 and RAD001 (Novartis, Basel, Switzerland) are now in clinical trials. Because of the known effect of rapamycin on lymphocyte proliferation, a potential problem with rapamycin analogues is immunosuppression. However, although prolonged immunosuppression can result from rapamycin and CCI-779 administered on continuous-dose schedules, the immunosuppressive effects of rapamycin analogues resolve in ~24 h after therapy (91). The principal toxicities of CCI-779 have included dermatological toxicity, myelosuppression, infection, mucositis, diarrhea, reversible elevations in liver function tests, hyperglycemia, hypokalemia, hypocalcemia, and depression (87, 92–94). Phase II trials of CCI-779 have been conducted in advanced renal cell carcinoma and in stage III/IV breast carcinoma patients who failed with prior chemotherapy. In the results reported in abstract form, although there were no complete responses, partial responses were documented in both renal cell carcinoma and in breast carcinoma (94, 95). Thus, CCI-779 has documented preliminary clinical activity in a previously treated, unselected patient population.

Active investigation is under way into patient selection for mTOR inhibitors. Several studies have found an enhanced efficacy of CCI-779 in PTEN-null tumors (30, 96). Another study found that six of eight breast cancer cell lines were responsive to CCI-779, although only two of these lines lacked PTEN (97). There was, however, a positive correlation between Akt activation and CCI-779 sensitivity (97). This correlation suggests that activation of the PI3K-Akt pathway,

regardless of whether it is attributable to a PTEN mutation or to overexpression of receptor tyrosine kinases, makes cancer cell amenable to mTOR-directed therapy. In contrast, lower levels of the target of mTOR, 4E-BP1, are associated with rapamycin resistance; thus, a lower 4E-BP1/eIF4E ratio may predict rapamycin resistance (98).

Another mode of activity for rapamycin and its analogues appears to be through inhibition of angiogenesis. This activity may be both through direct inhibition of endothelial cell proliferation as a result of mTOR inhibition in these cells or by inhibition of translation of such proangiogenic factors as vascular endothelial growth factor in tumor cells (99, 100).

The angiogenesis inhibitor tunicamycin, another anticancer drug currently under study, was also found recently to inhibit translation in endothelial cells (101). Through a requisite interaction with integrin, tunicamycin inhibits activation of the PI3K/Akt pathway and mTOR in endothelial cells and prevents dissociation of eIF4E from 4E-BP1, thereby inhibiting cap-dependent translation. These findings suggest that endothelial cells are especially sensitive to therapies targeting the mTOR-signaling pathway.

Targeting eIF2 α : EPA, Clotrimazole, mda-7, and Flavonoids

EPA is an n-3 polyunsaturated fatty acid found in the fish-based diets of populations having a low incidence of cancer (102). EPA inhibits the proliferation of cancer cells (103), as well as in animal models (104, 105). It blocks cell division by inhibiting translation initiation (105). EPA releases Ca²⁺ from intracellular stores while inhibiting their refilling, thereby activating PKR. PKR, in turn phosphorylates and inhibits eIF2 α , resulting in the inhibition of protein synthesis at the level of translation initiation. Similarly, clotrimazole, a potent antiproliferative agent *in vitro* and *in vivo*, inhibits cell growth through depletion of Ca²⁺ stores, activation of PKR, and phosphorylation of eIF2 α (106). Consequently, clotrimazole preferentially decreases the expression of cyclins A, E, and D1, resulting in blockage of the cell cycle in G₁.

mda-7 is a novel tumor suppressor gene being developed as a gene therapy agent. Adenoviral transfer of mda-7 (Ad-mda7) induces apoptosis in many cancer cells including breast, colorectal, and lung cancer (107–109). Ad-mda7 also induces and activates PKR, which leads to phosphorylation of eIF2 α and induction of apoptosis (110).

Flavonoids such as genistein and quercetin suppress tumor cell growth. All three mammalian eIF2 α kinases, PKR, heme-regulated inhibitor, and PERK/PEK, are activated by flavonoids, with phosphorylation of eIF2 α and inhibition of protein synthesis (111).

Targeting eIF4A and eIF4E: Antisense RNA and Peptides

Antisense expression of eIF4A decreases the proliferation rate of melanoma cells (112). Sequestration of eIF4E by overexpression of 4E-BP1 is proapoptotic and decreases tumorigenicity (113, 114). Reduction of eIF4E with antisense RNA decreases soft agar growth, increases tumor latency, and increases the rates of tumor doubling times (7). Antisense eIF4E RNA treat-

ment also reduces the expression of angiogenic factors (115) and has been proposed as a potential adjuvant therapy for head and neck cancers, particularly when elevated eIF4E is found in surgical margins. Small molecule inhibitors that bind the eIF4G/4E-BP1-binding domain of eIF4E are proapoptotic (116) and are also being actively pursued.

Exploiting Selective Translation for Gene Therapy

A different therapeutic approach that takes advantage of the enhanced cap-dependent translation in cancer cells is the use of gene therapy vectors encoding suicide genes with highly structured 5' UTR. These mRNA would thus be at a competitive disadvantage in normal cells and not translate well, whereas in cancer cells, they would translate more efficiently. For example, the introduction of the 5' UTR of fibroblast growth factor-2 5' to the coding sequence of *herpes simplex virus type-1 thymidine kinase* gene, allows for selective translation of *herpes simplex virus type-1 thymidine kinase* gene in breast cancer cell lines compared with normal mammary cell lines and results in selective sensitivity to ganciclovir (117).

Toward the Future

Translation is a crucial process in every cell. However, several alterations in translational control occur in cancer. Cancer cells appear to need an aberrantly activated translational state for survival, thus allowing the targeting of translation initiation with surprisingly low toxicity. Components of the translational machinery, such as eIF4E, and signal transduction pathways involved in translation initiation, such as mTOR, represent promising targets for cancer therapy. Inhibitors of the mTOR have already shown some preliminary activity in clinical trials. It is possible that with the development of better predictive markers and better patient selection, response rates to single-agent therapy can be improved. Similar to other cytostatic agents, however, mTOR inhibitors are most likely to achieve clinical utility in combination therapy. In the interim, our increasing understanding of translation initiation and signal transduction pathways promise to lead to the identification of new therapeutic targets in the near future.

Acknowledgments

We thank Gayle Nesom from The University of Texas M. D. Anderson Cancer Center Department of Scientific Publications for editorial assistance and Dr. Elmer Bernstein for assistance with manuscript preparation.

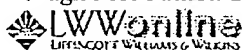
References

1. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. USA*, 98: 7029–7036, 2001.
2. Rosenwald, I. B., Kaspar, R., Rousseau, D., Gehlke, L., Leboucq, P., Chen, J. J., Schmidt, E. V., Sonenberg, N., and London, I. M. Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. *J. Biol. Chem.*, 270: 21176–21180, 1995.
3. Rosenwald, I. B., Hutzler, M. J., Wang, S., Savas, L., and Fraire, A. E. Expression of eukaryotic translation initiation factors 4E and 2 α is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. *Cancer (Phila.)*, 92: 2164–2171, 2001.

4. Darveau, A., Pelletier, J., and Sonenberg, N. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. USA*, 82: 2315-2319, 1985.
5. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA*, 83: 2850-2854, 1986.
6. Koromilas, A. E., Lazaris-Karatzas, A., and Sonenberg, N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.*, 11: 4153-4158, 1992.
7. Rinker-Schaeffer, C. W., Graff, J. R., De Benedetti, A., Zimmer, S. G., and Rhoads, R. E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. *Int. J. Cancer*, 55: 841-847, 1993.
8. Kevil, C. G., De Benedetti, A., Payne, D. K., Coe, L. L., Laroux, F. S., and Alexander, J. S. Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. *Int. J. Cancer*, 65: 785-790, 1996.
9. Nathan, C. A., Franklin, S., Abreo, F. W., Nassar, R., De Benedetti, A., and Glass, J. Analysis of surgical margins with the molecular marker eIF4E: a prognostic factor in patients with head and neck cancer. *J. Clin. Oncol.*, 17: 2909-2914, 1999.
10. Fukunaga, R., and Hunter, T. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *EMBO J.*, 16: 1921-1933, 1997.
11. Waskiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. *EMBO J.*, 16: 1909-1920, 1997.
12. Wang, X., Flynn, A., Waskiewicz, A. J., Webb, B. L., Vries, R. G., Baines, I. A., Cooper, J. A., and Proud, C. G. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. *J. Biol. Chem.*, 273: 9373-9377, 1998.
13. Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO J.*, 18: 270-279, 1999.
14. Kleijn, M., Schepers, G. C., Voorma, H. O., and Thomas, A. A. Regulation of translation initiation factors by signal transduction. *Eur. J. Biochem.*, 253: 531-544, 1998.
15. Raught, B., and Gingras, A. C. eIF4E activity is regulated at multiple levels. *Int. J. Biochem. Cell Biol.*, 31: 43-57, 1999.
16. Takeuchi, K., Shibamoto, S., Nagamine, K., Shigemori, I., Omura, S., Kitamura, N., and Ito, F. Signaling pathways leading to transcription and translation cooperatively regulate the transient increase in expression of c-Fos protein. *J. Biol. Chem.*, 276: 26077-26083, 2001.
17. Kawasome, H., Papst, P., Webb, S., Keller, G. M., Johnson, G. L., Gelland, E. W., and Terada, N. Targeted disruption of p70(S6K) defines its role in protein synthesis and rapamycin sensitivity. *Proc. Natl. Acad. Sci. USA*, 95: 5033-5038, 1998.
18. Christle, G. R., Hajdich, E., Hundal, H. S., Proud, C. G., and Taylor, P. M. Intracellular sensing of amino acids in *Xenopus laevis* oocytes stimulates p70 S6 kinase in a target of rapamycin-dependent manner. *J. Biol. Chem.*, 277: 9952-9957, 2002.
19. Hara, K., Yonezawa, K., Weng, Q. P., Kozlowski, M. T., Belham, C., and Avruch, J. Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.*, 273: 14484-14494, 1998.
20. Graves, L. M., Bornfeldt, K. E., Argast, G. M., Krebs, E. G., Kong, X., Lin, T. A., and Lawrence, J. C., Jr. cAMP- and rapamycin-sensitive regulation of the association of eukaryotic initiation factor 4E and the translational regulator PHAS-I in aortic smooth muscle cells. *Proc. Natl. Acad. Sci. USA*, 92: 7222-7226, 1995.
21. Merrick, W. C., and Hershey, J. W. B. The pathway and mechanism of eukaryotic protein synthesis. In: J. W. B. Hershey and M. B. Mathews (eds.), *Translational Control*, pp. 31-69. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1996.
22. Kimball, S. R. Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.*, 31: 25-29, 1999.
23. Jagus, R., Joshi, B., and Barber, G. N. PKR, apoptosis and cancer. *Int. J. Biochem. Cell Biol.*, 31: 123-138, 1999.
24. Thomas, G., and Hall, M. N. TOR signalling and control of cell growth. *Curr. Opin. Cell Biol.*, 9: 782-787, 1997.
25. Gingras, A. C., Raught, B., and Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev.*, 15: 807-826, 2001.
26. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R., and Sonenberg, N. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.*, 13: 1422-1437, 1999.
27. Kumar, V., Pandey, P., Sabatini, D., Kumar, M., Majumder, P. K., Bharti, A., Carmichael, G., Kufe, D., and Karhanda, S. Functional interaction between RAFT1/FRAP/mTOR and protein kinase C δ in the regulation of cap-dependent initiation of translation. *EMBO J.*, 19: 1087-1097, 2000.
28. Yang, D. Q., and Kastan, M. B. Participation of ATM in insulin signaling through phosphorylation of eIF-4E-binding protein 1. *Nat. Cell Biol.*, 2: 893-898, 2000.
29. Liu, G., Zhang, Y., Bode, A. M., Ma, W. Y., and Dong, Z. Phosphorylation of 4E-BP1 is mediated by the p38/MSK1 pathway in response to UVB irradiation. *J. Biol. Chem.*, 277: 8810-8816, 2002.
30. Neshat, M. S., Mellinghoff, I. K., Tran, C., Stiles, B., Thomas, G., Petersen, R., Frost, P., Gibbons, J. J., Wu, H., and Sawyers, C. L. Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. *Proc. Natl. Acad. Sci. USA*, 98: 10314-10319, 2001.
31. Sekulic, A., Hudson, C. C., Homme, J. L., Yin, P., Otemess, D. M., Kamitz, L. M., and Abraham, R. T. A direct linkage between the phosphoinositide 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. *Cancer Res.*, 60: 3504-3513, 2000.
32. Scott, P. H., and Lawrence, J. C., Jr. Attenuation of mammalian target of rapamycin activity by increased cAMP in 3T3-L1 adipocytes. *J. Biol. Chem.*, 273: 34496-34501, 1998.
33. Reynolds, I. T., Bodine, S. C., and Lawrence, J. C., Jr. Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. *J. Biol. Chem.*, 277: 17657-17662, 2002.
34. Peterson, R. T., Beal, P. A., Comb, M. J., and Schreiber, S. L. FKBP12-rapamycin-associated protein (FRAP) autophosphorylates at serine 2481 under translationally repressive conditions. *J. Biol. Chem.*, 275: 7416-7423, 2000.
35. Peterson, R. T., Desai, B. N., Hardwick, J. S., and Schreiber, S. L. Protein phosphatase 2A interacts with the 70-kDa S6 kinase and is activated by inhibition of FKBP12-rapamycin-associated protein. *Proc. Natl. Acad. Sci. USA*, 96: 4438-4442, 1999.
36. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, 3: 803-815, 1989.
37. Sheets, M. D., Wu, M., and Wickens, M. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature (Lond.)*, 374: 511-516, 1995.
38. Varnum, S. M., and Womington, W. M. Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific cis-sequences: a default mechanism for translational control. *Genes Dev.*, 4: 2278-2286, 1990.
39. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.*, 5: 2108-2116, 1991.
40. Sachs, A. B., and Varani, G. Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat. Struct. Biol.*, 7: 356-361, 2000.
41. Wolffe, A. P., and Meric, F. Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int. J. Biochem. Cell Biol.*, 28: 247-257, 1996.
42. Evdokimova, V. M., Wei, C. L., Sitkov, A. S., Simonenko, P. N., Lazarev, O. A., Vasilenko, K. S., Ustinov, V. A., Hershey, J. W., and Ovchinnikov, L. P. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. *J. Biol. Chem.*, 270: 3186-3192, 1995.
43. Matsumoto, K., Meric, F., and Wolffe, A. P. Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. *J. Biol. Chem.*, 271: 22706-22712, 1996.

44. Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L. P., and Sonenberg, N. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *EMBO J.*, 20: 5491-5502, 2001.
45. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Toneygawa, S. Activation of the *c-myc* gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. USA*, 80: 7476-7480, 1983.
46. Nanbru, C., Lafon, I., Audigier, S., Gensac, M. C., Vagner, S., Huez, G., and Prats, A. C. Alternative translation of the proto-oncogene *c-myc* by an internal ribosome entry site. *J. Biol. Chem.*, 272: 32061-32066, 1997.
47. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A., and Willis, A. E. *c-Myc* 5' untranslated region contains an internal ribosome entry segment. *Oncogene*, 16: 423-428, 1998.
48. Paulin, F. E., West, M. J., Sullivan, N. F., Whitney, R. L., Lyne, L., and Willis, A. E. Aberrant translational control of the *c-myc* gene in multiple myeloma. *Oncogene*, 13: 505-513, 1996.
49. Chappell, S. A., LeQuesne, J. P., Paulin, F. E., de Schoolmeester, M. L., Stoneley, M., Soutar, R. L., Raiston, S. H., Hellrich, M. H., and Willis, A. E. A mutation in the *c-myc*-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene*, 19: 4437-4440, 2000.
50. Signori, E., Bagni, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F., and Fazio, V. M. A somatic mutation in the 5'UTR of *BRCA1* gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene*, 20: 4596-4600, 2001.
51. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.*, 21: 128-132, 1999.
52. Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Sarti, A., Rosenthal, A., Etroy-Stein, O., Shih, Y., and Rotman, G. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of *ATM* gene expression. *Nucleic Acids Res.*, 25: 1678-1684, 1997.
53. Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., and Morris, D. R. Role of two upstream open reading frames in the translational control of oncogene *mdm2*. *Oncogene*, 18: 5631-5637, 1999.
54. Sobczak, K., and Krzyzosiak, W. J. Structural determinants of *BRCA1* translational regulation. *J. Biol. Chem.*, 277: 17349-17358, 2002.
55. Arrick, B. A., Grendell, R. L., and Griffin, L. A. Enhanced translational efficiency of a novel transforming growth factor β mRNA in human breast cancer cells. *Mol. Cell. Biol.*, 14: 619-628, 1994.
56. Myers, S. M., Eng, C., Ponder, B. A., and Mulligan, L. M. Characterization of *RET* proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for *RET*. *Oncogene*, 11: 2039-2045, 1995.
57. Byrne, J. A., Tomasetto, C., Rouyer, N., Bellocq, J. P., Rio, M. C., and Basset, P. The tissue inhibitor of metalloproteinases-3 gene in breast carcinoma: identification of multiple polyadenylation sites and a stromal pattern of expression. *Mol. Med.*, 1: 418-427, 1995.
58. Moscow, J. A., He, R., Gudas, J. M., and Cowan, K. H. Utilization of multiple polyadenylation signals in the human *RHOA* protooncogene. *Gene (Amst.)*, 144: 229-236, 1994.
59. Senterre-Lesenfants, S., Alag, A. S., and Sobel, M. E. Multiple mRNA species are generated by alternate polyadenylation from the human *calmodulin-1* gene. *J. Cell. Biochem.*, 58: 445-454, 1995.
60. Lazaris-Karatzas, A., Montine, K. S., and Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature (Lond.)*, 345: 544-547, 1990.
61. De Benedetti, A., and Rhoads, R. E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. *Proc. Natl. Acad. Sci. USA*, 87: 8212-8216, 1990.
62. Polunovsky, V. A., Rosenwald, I. B., Tan, A. T., White, J., Chiang, L., Sonenberg, N., and Bitterman, P. B. Translational control of programmed cell death: eukaryotic translation initiation factor 4E blocks apoptosis in growth-factor-restricted fibroblasts with physiologically expressed or de-regulated *Myc*. *Mol. Cell. Biol.*, 16: 6573-6581, 1996.
63. Miyagi, Y., Sugiyama, A., Asai, A., Okazaki, T., Kuchino, Y., and Kerr, S. J. Elevated levels of eukaryotic translation initiation factor eIF-4E mRNA in a broad spectrum of transformed cell lines. *Cancer Lett.*, 97: 247-252, 1995.
64. Kerekatte, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas. *Int. J. Cancer*, 64: 27-31, 1995.
65. Li, B. D., Liu, L., Dawson, M., and De Benedetti, A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. *Cancer (Phila.)*, 79: 2385-2390, 1997.
66. Sorrells, D. L., Black, D. R., Meschonat, C., Rhoads, R., De Benedetti, A., Gao, M., Williams, B. J., and Li, B. D. Detection of eIF4E gene amplification in breast cancer by competitive PCR. *Ann. Surg. Oncol.*, 5: 232-237, 1998.
67. Li, B. D., McDonald, J. C., Nassar, R., and De Benedetti, A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. *Ann. Surg.*, 227: 756-761; discussion, 761-763, 1998.
68. Wang, S., Rosenwald, I. B., Hutzler, M. J., Phan, G. A., Savas, L., Chen, J. J., and Woda, B. A. Expression of the eukaryotic translation initiation factors 4E and 2a in non-Hodgkin's lymphomas. *Am. J. Pathol.*, 155: 247-255, 1999.
69. Fukuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Ekimoto, H., and Igarashi, K. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.*, 57: 5041-5044, 1997.
70. Eberle, J., Krasagakis, K., and Orfanos, C. E. Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells *in vitro*. *Int. J. Cancer*, 71: 396-401, 1997.
71. Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Wakatsuki, T., Hada, A., Goseki, N., Igarashi, T., Hattuse, K., Aihara, T., Horuchi, S., Shichita, M., Yamamoto, N., and Yamamoto, M. Enhanced expression of translation factor mRNAs in hepatocellular carcinoma. *Anticancer Res.*, 20: 2489-2494, 2000.
72. Nupponen, N. N., Porkka, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J., and Visakorpi, T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. *Am. J. Pathol.*, 154: 1777-1783, 1999.
73. Rothe, M., Ko, Y., Albers, P., and Wernert, N. Eukaryotic initiation factor 3 p110 mRNA is overexpressed in testicular seminomas. *Am. J. Pathol.*, 157: 1597-1604, 2000.
74. Barklund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.* (Bethesda), 92: 1252-1259, 2000.
75. Topalian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. *Mol. Cell. Biol.*, 21: 5614-5623, 2001.
76. Scorilas, A., Tafari, M., Aravanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsalpakis, C. M., and Trangas, T. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: a new independent prognostic marker in primary breast cancer. *Cancer Res.*, 60: 5427-5433, 2000.
77. Janz, M., Harbeck, N., Dettmar, P., Berger, U., Schmidt, A., Jurchott, K., Schmitt, M., and Royer, H. D. Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. *Int. J. Cancer*, 97: 278-282, 2002.
78. Shibahara, K., Sugio, K., Osaki, T., Uchiyama, T., Maehara, Y., Kohno, K., Yasumoto, K., Sugimachi, K., and Kuwano, M. Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. *Clin. Cancer Res.*, 7: 3151-3155, 2001.
79. Kamura, T., Yahata, H., Amada, S., Ogawa, S., Sonoda, T., Kobayashi, H., Mitsumoto, M., Kohno, K., Kuwano, M., and Nakano, H. Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? *Cancer (Phila.)*, 85: 2450-2454, 1999.
80. Bargou, R. C., Jurchott, K., Wagener, C., Bergmann, S., Metzner, S., Bonnmert, K., Mapara, M. Y., Winzer, K. J., Dietel, M., Dorken, B., and Royer, H. D. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic *MDR1* gene expression. *Nat. Med.*, 3: 447-450, 1997.
81. Aoki, M., Blazek, E., and Vogt, P. K. A role of the kinase mTOR in cellular transformation induced by the oncoproteins P3k and Akt. *Proc. Natl. Acad. Sci. USA*, 98: 136-141, 2001.
82. Child, S. J., Miller, M. K., and Geballe, A. P. Cell type-dependent and -independent control of HER-2/neu translation. *Int. J. Biochem. Cell Biol.*, 31: 201-213, 1999.

83. Jefferies, H. B., Reinhard, C., Kozma, S. C., and Thomas, G. Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. *Proc. Natl. Acad. Sci. USA*, 91: 4441-4445, 1994.
84. Terada, N., Patel, H. R., Takase, K., Kohno, K., Naim, A. C., and Gelfand, E. W. Rapamycin selectively inhibits translation of mRNAs encoding elongation factors and ribosomal proteins. *Proc. Natl. Acad. Sci. USA*, 91: 11477-11481, 1994.
85. Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. *EMBO J.*, 16: 3693-3704, 1997.
86. Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N., and Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. *EMBO J.*, 15: 658-664, 1996.
87. Hidalgo, M., and Rowinsky, E. K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. *Oncogene*, 19: 6680-6686, 2000.
88. Hosoi, H., Dilling, M. B., Shikata, T., Liu, L. N., Shu, L., Ashmun, R. A., Germain, G. S., Abraham, R. T., and Houghton, P. J. Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. *Cancer Res.*, 59: 886-894, 1999.
89. Huang, S., and Houghton, P. J. Resistance to rapamycin: a novel anticancer drug. *Cancer Metastasis Rev.*, 20: 69-78, 2001.
90. Goegerger, B., Kerr, K., Tang, C. B., Fung, K. M., Powell, B., Sutton, L. N., Phillips, P. C., and Janss, A. J. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/medulloblastoma models as single agent and in combination chemotherapy. *Cancer Res.*, 61: 1527-1532, 2001.
91. Gibbons, J. J., D'Iscafi, C., Peterson, R., Hernandez, R., Skotnicki, J., and Frost, P. The effect of CCI-779, a novel macrolide anti-tumor agent, on the growth of human tumor cells *in vitro* and in nude mouse xenografts *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 40: 301, 1999.
92. Hidalgo, M., Rowinsky, E., Erickman, C., Marshall, B., Marks, R., Edwards, T., and Duckner, J. A Phase I and pharmacological study of CCI-779 cycle inhibitor. *Ann. Oncol.*, 11 (Suppl. 4): 133, 2001.
93. Alexandre, J., Raymond, E., Depenbrock, H., Mekhaldi, S., Angevin, E., Paillet, C., Hanauske, A., Frisch, J., Feussner, A., and Armand, J. P. CCI-779, a new rapamycin analog, has antitumor activity at doses inducing only mild cutaneous effects and mucositis: early results of an ongoing Phase I study. *Proceedings of the 1999 AACR-NCI-EORTC International Conference, Clin. Cancer Res.*, 5 (Suppl.): 3730s, 1999.
94. Chan, S., Johnston, S., Scheulen, M. E., Mross, K., Morant, A., Lahr, A., Feussner, A., Berger, M., and Kirsch, T. First report: a Phase 2 study of the safety and activity of CCI-779 for patients with locally advanced or metastatic breast cancer failing prior chemotherapy. *Proc. Am. Soc. Clin. Oncol.*, 21: 44a, 2002.
95. Atkins, M. B., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J. P., Hudes, G., Park, Y., Marshall, B., Boni, J., and Dulciet, G. A randomized double-blind Phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. *Proc. Am. Soc. Clin. Oncol.*, 21: 10a, 2002.
96. Smith, S. G., Trinh, C. M., Inge, L. J., Thomas, G., Cloughesy, T. F., Sawyers, C. L., and Mischel, P. S. PTEN expression status predicts glioblastoma cell sensitivity to CCI-779. *Proc. Am. Assoc. Cancer Res.*, 43: 335, 2002.
97. Yu, K., Toral-Barza, L., D'Iscafi, C., Zhang, W. G., Skotnicki, J., Frost, P., and Gibbons, J. J. mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. *Endocr. Relat. Cancer*, 8: 249-258, 2001.
98. Dilling, M. B., Germain, G. S., Dudkin, L., Jayaraman, A. L., Zhang, X., Harwood, F. C., and Houghton, P. J. 4E-binding proteins, the suppressors of eukaryotic initiation factor 4E, are downregulated in cells with acquired or intrinsic resistance to rapamycin. *J. Biol. Chem.*, 277: 13907-13917, 2002.
99. Guba, M., von Breitenbuch, P., Steinbauer, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., and Geissler, E. K. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. *Nat. Med.*, 8: 128-135, 2002.
100. Lane, H. A., Schell, C., Theuer, A., O'Reilly, T., and Wood, J. Anti-angiogenic activity of RAD001, an orally active anticancer agent. *Proc. Am. Assoc. Cancer Res.*, 43: 184, 2002.
101. Maeshima, Y., Sudhakar, A., Lively, J. C., Ueki, K., Kharbada, S., Kahn, C. R., Sonenberg, N., Hynes, R. O., and Kalluri, R. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. *Science (Wash. DC)*, 295: 140-143, 2002.
102. Caygill, C. P., Charlett, A., and Hill, M. J. Fat, fish oil and cancer. *Br. J. Cancer*, 74: 159-164, 1996.
103. Falconer, J. S., Ross, J. A., Fearon, K. C., Hawkins, R. A., O'Riordan, M. G., and Carter, D. C. Effect of eicosapentaenoic acid and other fatty acids on the growth *in vitro* of human pancreatic cancer cell lines. *Br. J. Cancer*, 69: 826-832, 1994.
104. Noguchi, M., Minami, M., Yagasaki, R., Kinoshita, K., Earashi, M., Kitagawa, H., Taniya, T., and Miyazaki, I. Chemoprevention of DMBA-induced mammary carcinogenesis in rats by low-dose EPA and DHA. *Br. J. Cancer*, 75: 348-353, 1997.
105. Palakurthi, S. S., Fluckiger, R., Aktas, H., Changolkar, A. K., Shahsafari, A., Hameit, S., Klic, E., and Halperin, J. A. Inhibition of translation initiation mediates the anticancer effect of the n-3 polyunsaturated fatty acid eicosapentaenoic acid. *Cancer Res.*, 60: 2919-2925, 2000.
106. Aktas, H., Fluckiger, R., Acosta, J. A., Savage, J. M., Palakurthi, S. S., and Halperin, J. A. Depletion of intracellular Ca^{2+} stores, phosphorylation of eIF2 α , and sustained inhibition of translation initiation mediate the anticancer effects of clofazimine. *Proc. Natl. Acad. Sci. USA*, 95: 8280-8285, 1998.
107. Mhashilkar, A. M., Schrock, R. D., Hindi, M., Liao, J., Sieger, K., Kourouna, F., Zou-Yang, X. H., Onishi, E., Tak, O., Vedvick, T. S., Fanger, G., Stewart, L., Watson, G. J., Snary, D., Fisher, P. B., Saeki, T., Roth, J. A., Ramesh, R., and Chada, S. Melanoma differentiation associated gene-7 (*mda-7*): a novel anti-tumor gene for cancer gene therapy. *Mol. Med.*, 7: 271-282, 2001.
108. Su, Z. Z., Madireddi, M. T., Lin, J. J., Young, C. S., Kitada, S., Reed, J. C., Goldstein, N. I., and Fisher, P. B. The cancer growth suppressor gene *mda-7* selectively induces apoptosis in human breast cancer cells and inhibits tumor growth in nude mice. *Proc. Natl. Acad. Sci. USA*, 95: 14400-14405, 1998.
109. Saeki, T., Mhashilkar, A., Chada, S., Branch, C., Roth, J. A., and Ramesh, R. Tumor-suppressive effects by adenovirus-mediated *mda-7* gene transfer in non-small cell lung cancer cell *in vitro*. *Gene Ther.*, 7: 2051-2057, 2000.
110. Pataer, A., Vorburger, S. A., Barber, G. N., Chada, S., Mhashilkar, A. M., Zou-Yang, H., Stewart, A. L., Balachandran, S., Roth, J. A., Hunt, K. K., and Swisher, S. G. Adenoviral transfer of the melanoma differentiation-associated gene 7 (*mda-7*) induces apoptosis of lung cancer cells via up-regulation of the double-stranded RNA-dependent protein kinase (PKR). *Cancer Res.*, 62: 2239-2243, 2002.
111. Ito, T., Wamken, S. P., and May, W. S. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2 α kinases. *Biochem. Biophys. Res. Commun.*, 265: 589-594, 1999.
112. Eberle, J., Fecker, L. F., Bittner, J. U., Orlanos, C. E., and Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. *Br. J. Cancer*, 86: 1957-1962, 2002.
113. Polunovsky, V. A., Gingras, A. C., Sonenberg, N., Peterson, M., Tan, A., Rubins, J. B., Manivel, J. C., and Bitterman, P. B. Translational control of the antiapoptotic function of Ras. *J. Biol. Chem.*, 275: 24776-24780, 2000.
114. D'Cunha, J., Kratzke, M. G., Alter, M. D., Polunovsky, V. A., Bitterman, P. B., and Kratzke, R. A. Over-expression of the translational repressor 4E-BP1 inhibits NSCLC tumorigenicity *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 43: 816-817, 2002.
115. DeFatta, R. J., Nathan, C. A., and De Benedetti, A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. *Laryngoscope*, 110: 928-933, 2000.
116. Herbert, T. P., Fahraeus, R., Prescott, A., Lane, D. P., and Proud, C. G. Rapid induction of apoptosis mediated by peptides that bind initiation factor eIF4E. *Curr. Biol.*, 10: 793-795, 2000.
117. DeFatta, R. J., Li, Y., and De Benedetti, A. Selective killing of cancer cells based on translational control of a suicide gene. *Cancer Gene Ther.*, 9: 573-578, 2002.



Correlative immunohistochemical and reverse transcriptase polymerase chain reaction analysis of somatostatin receptor type 2 in neuroendocrine tumors of the lung.

Papotti M, Croce S, Macri L, Funaro A, Pecchioni C, Schindler M, Bussolati G.

Department of Biomedical Sciences and Oncology, University of Turin, Italy.

Somatostatin receptors type 2 (sst2) have been frequently detected in neuroendocrine tumors and bind somatostatin analogues, such as octreotide, with high affinity. Receptor autoradiography, specific mRNA detection and, more recently, antisst2 polyclonal antibodies are currently employed to reveal sst2. The aim of the present study was to investigate by three different techniques the presence of sst2 in a series of 26 neuroendocrine tumors of the lung in which fresh frozen tissue and paraffin sections were available. It was possible, therefore, to compare, in individual cases, RNA analysis studied by reverse transcriptase polymerase chain reaction (RT-PCR), in situ hybridization (ISH), and immunohistochemistry. A series of 20 nonneuroendocrine lung carcinoma samples served as controls. RT-PCR was positive for sst2 in 22 of 26 samples, including 15 of 15 typical carcinoids, 5 of 6 atypical carcinoids, and 2 of 5 small-cell carcinomas. The sst2 mRNA signal obtained by RT-PCR was strong in the majority (87%) of typical carcinoids and of variable intensity in atypical carcinoids and small-cell carcinomas. A weakly positive signal was observed in 5 of 20 control samples. In immunohistochemistry, two different antibodies (anti-sst2) were employed, including a monoclonal antibody, generated in the Department of Pathology, University of Turin. In the majority of samples a good correlation between sst2 mRNA (as detected by RT-PCR) and sst2 protein expression (as detected by immunohistochemistry) was observed. However, one atypical carcinoid and one small-cell carcinoma had focal immunostaining but no RT-PCR signal. ISH performed in selected samples paralleled the results obtained with the other techniques. A low sst2 expression was associated with high grade neuroendocrine tumors and with aggressive behavior. It is concluded that 1) neuroendocrine tumors of the lung express sst2, and there is a correlation between the mRNA amount and the degree of differentiation; 2) immunohistochemistry and ISH are reliable tools to demonstrate sst2 in these tumors; and 3) sst2 identification in tissue sections may provide information on the diagnostic or therapeutic usefulness of somatostatin analogues in individual patients with neuroendocrine tumors.

PMID: 10718213 [PubMed - indexed for MEDLINE]



Expression of somatostatin receptor types 1-5 in 81 cases of gastrointestinal and pancreatic endocrine tumors. A correlative immunohistochemical and reverse-transcriptase polymerase chain reaction analysis.

Papotti M, Bongiovanni M, Volante M, Allia E, Landolfi S, Helboe L, Schindler M, Cole SL, Bussolati G.

Department of Biomedical Sciences and Oncology, University of Turin, Via Santena 7, 10126 Turin, Italy. mauro.papotti@unito.it

Somatostatin receptors (SSTRs) have been extensively mapped in human tumors by means of autoradiography, reverse-transcriptase polymerase chain reaction (RT-PCR), in situ hybridization (ISH) and immunohistochemistry (IHC). We analyzed the SSTR type 1-5 expression by means of RT-PCR and/or IHC in a series of 81 functioning and non-functioning gastroenteropancreatic (GEP) endocrine tumors and related normal tissues. Moreover, we compared the results with clinical, pathological and hormonal features. Forty-six cases (13 intestinal and 33 pancreatic) were studied for SSTR 1-5 expression using RT-PCR, IHC with antibodies to SSTR types 2, 3, 5 and ISH for SSTR2 mRNA. The vast majority of tumors expressed SSTR types 1, 2, 3 and 5, while SSTR4 was detected in a small minority. Due to the good correlation between RT-PCR and IHC data on SSTR types 2, 3, and 5, thirty-five additional GEP endocrine tumors were studied with IHC alone. Pancreatic insulinomas had an heterogeneous SSTR expression, while 100% of somatostatinomas expressed SSTR5 and 100% gastrinomas and glucagonomas expressed SSTR2. Pre-operative biopsy material showed an overlapping immunoreactivity with that of surgical specimens, suggesting that the SSTR status can be detected in the diagnostic work-up. It is concluded that SSTRs 1-5 are heterogeneously expressed in GEP endocrine tumors and that IHC is a reliable tool to detect SSTR types 2, 3 and 5 in surgical and biopsy specimens.

PMID: 12021920 [PubMed - indexed for MEDLINE]

MetaPress

GLUT1 messenger RNA and protein induction relates to the malignant transformation of cervical cancer.

Rudlowski C, Becker AJ, Schroder W, Rath W, Buttner R, Moser M.

Dept of Gynecology and Obstetrics, University Hospital Heidelberg, Vossstr 7-9, D-69115 Heidelberg, Germany.

We studied whether induction of glucose transporters (GLUTs) 1 to 4 correlates with human papillomavirus (HPV)-dependent malignant transformation of cervical epithelium. Tissue samples of cervical intraepithelial neoplasia (CIN; grades 1 to 3), invasive carcinomas, and lymph node metastasis were examined. HPV typing was performed. Tissue sections were immunostained with GLUT1 to GLUT4 antibodies. Messenger RNA (mRNA) in situ hybridization confirmed GLUT1 protein expression. Weak expression of GLUT1 was found in nondysplastic HPV-positive and HPV-negative epithelium; significant expression was observed in preneoplastic lesions, correlating with the degree of dysplasia. In CIN 3 high-risk HPV lesions, cervical cancer, and metastasis, GLUT1 was expressed at highest levels with a strong correlation of GLUT1 mRNA and protein expression. Immunostains for GLUT2 to GLUT4 were negative. Cervical tumor cells respond to enhanced glucose utilization by up-regulation of GLUT1. The strong induction of GLUT1 mRNA and protein in HPV-positive CIN 3 lesions suggests GLUT1 overexpression as an early event in cervical neoplasia. GLUT1 is potentially relevant as a diagnostic tool and glucose metabolism as a therapeutic target in cervical cancer.

PMID: 14608894 [PubMed - indexed for MEDLINE]

Virchows Arch (2002) 440:461–475
DOI 10.1007/s00428-002-0609-x

ORIGINAL ARTICLE

M. Papotti · M. Bongiovanni · M. Volante · E. Allia
S. Landolfi · L. Helboe · M. Schindler · S.L. Cole
G. Bussolati

Expression of somatostatin receptor types 1–5 in 81 cases of gastrointestinal and pancreatic endocrine tumors

A correlative immunohistochemical and reverse-transcriptase polymerase chain reaction analysis

Received: 2 August 2001 / Accepted: 21 December 2001 / Published online: 23 March 2002
© Springer-Verlag 2002

Abstract Somatostatin receptors (SSTRs) have been extensively mapped in human tumors by means of autoradiography, reverse-transcriptase polymerase chain reaction (RT-PCR), in situ hybridization (ISH) and immunohistochemistry (IHC). We analyzed the SSTR type 1–5 expression by means of RT-PCR and/or IHC in a series of 81 functioning and non-functioning gastroenteropancreatic (GEP) endocrine tumors and related normal tissues. Moreover, we compared the results with clinical, pathological and hormonal features. Forty-six cases (13 intestinal and 33 pancreatic) were studied for SSTR 1–5 expression using RT-PCR, IHC with antibodies to SSTR types 2, 3, 5 and ISH for SSTR2 mRNA. The vast majority of tumors expressed SSTR types 1, 2, 3 and 5, while SSTR4 was detected in a small minority. Due to the good correlation between RT-PCR and IHC data on SSTR types 2, 3, and 5, thirty-five additional GEP endocrine tumors were studied with IHC alone. Pancreatic insulinomas had an heterogeneous SSTR expression, while 100% of somatostatinomas expressed SSTR5 and 100% gastrinomas and glucagonomas expressed SSTR2. Pre-operative biopsy material showed an overlapping immunoreactivity with that of surgical specimens, suggesting that the SSTR status can be detected in the diagnostic work-

up. It is concluded that SSTRs 1–5 are heterogeneously expressed in GEP endocrine tumors and that IHC is a reliable tool to detect SSTR types 2, 3 and 5 in surgical and biopsy specimens.

Keywords GEP · Somatostatin receptors · Pancreatic islets · Neuroendocrine tumor · Immunohistochemistry · RT-PCR

Introduction

The role of somatostatin analogues in the clinical setting is well established: long-acting analogues have been used for more than 10 years to treat endocrine tumors, and octreotide scintigraphy is widely used for tumor localization and even for radio-guided surgery [5, 9, 15, 16, 19, 22, 45]. The rationale behind these clinical applications is based on the widespread occurrence of somatostatin receptors (here abbreviated SSTR – according to the GenBank nomenclature) in human tumors, with special reference to endocrine tumors [42]. To date, five SSTR subtypes have been identified (types 1–5) [23, 27, 50, 51] with different affinities for the natural ligand somatostatin as well as for synthetic analogues. SSTRs have been extensively mapped in human endocrine and non-endocrine tumors by means of autoradiography, Northern blot, reverse-transcriptase polymerase chain reaction (RT-PCR) and in situ hybridization (ISH) [4, 10, 14, 20, 24, 29, 30, 31, 32, 33, 34, 39, 44]. Irrespective of the original location of the tumor, most studies demonstrated a high incidence of SSTR1 and SSTR2 (and to a minor degree of the other types).

In recent years, several authors have developed antibodies for immunohistochemical demonstration of SSTR subtypes [8, 13, 17, 35, 40, 41, 43]. Such polyclonal antibodies, raised against peptide fragments of all five known SSTR subtypes, were mostly used for the immu-

M. Papotti (✉) · M. Bongiovanni · M. Volante · E. Allia
S. Landolfi · G. Bussolati
Department of Biomedical Sciences and Oncology,
University of Turin, Via Santena 7, 10126 Torino, Italy
e-mail: mauro.papotti@unito.it
Tel.: +39-011-6706514, Fax: +39-011-6635267

L. Helboe
Lundbeck, Copenhagen, Denmark

M. Schindler · S.L. Cole
Glaxo Institute of Applied Pharmacology,
University of Cambridge, UK

M. Schindler
Boehringer Ingelheim Pharma KG, Cardiovascular research,
Biberach, Germany

nohistochemical demonstration of SSTR in brain regions or in normal human or animal tissues. RT-PCR has the great advantage of a high sensitivity, but does not allow the precise SSTR cellular localization, which can be obtained by parallel ISH or immunohistochemical investigations [17, 25, 32, 33, 43].

In gastro-entero-pancreatic (GEP) endocrine tumors, SSTRs were analyzed using various procedures, including immunohistochemistry (IHC) for SSTR2. Expression of this receptor type in a variable percentage of pancreatic tumors and intestinal carcinoids, and a correlation between presence of SSTRs and scintigraphic localization and/or response to octreotide therapy were reported [13, 19, 22]. However, individual methods to reveal SSTRs were compared and correlated in a limited case series of GEP endocrine tumors, and immunohistochemical detection was restricted to SSTR2A analysis only. An exception was the large studies of the groups of Jais and Wulbrand [12, 49], who investigated all five SSTR subtypes using the highly sensitive RT-PCR technique and found high levels of SSTR1, 2 and 5 mRNA expression and low amounts of SSTR3 and SSTR4 mRNA. However, no parallel data on SSTR cellular localization were presented.

The aim of this study was to analyze the expression of SSTRs 1-5 by means of RT-PCR and/or IHC in a large series of 81 functioning and non-functioning GEP endocrine tumors and to compare the results with clinical, pathological and hormonal features. Specific antibodies to SSTRs 2, 3 and 5 (which are the subtypes with the highest affinity for octreotide) were used in parallel with the analysis of mRNA expression levels. A heterogeneous distribution of SSTR in both pancreatic and intestinal endocrine tumors was observed and the immunohistochemical SSTR detection was found to be a reliable diagnostic tool in both surgical and pre-operative biopsy specimens, relative to mRNA analysis.

Materials and methods

Case selection

Forty-six GEP endocrine tumors, in which fresh frozen material was available for mRNA analysis (13 and 33 tumors of gastrointestinal and pancreatic origin, respectively), were collected from the pathology files of the University of Turin. These tumors were studied for SSTR expression by means of RT-PCR and IHC. Since the results of the two procedures showed a good correlation (see below), the study was expanded to 35 additional GEP neoplasms in which formalin-fixed and paraffin-embedded tissue blocks were available for immunohistochemical staining only. The reason for entering additional cases was to investigate SSTR expression in all types of hormone-producing pancreatic tumors and to analyze a series of aggressive (metastatic) intestinal carcinoids to compare SSTR expression with the tumor stage. Overall, 81 cases were entered in the present study, including 28 gastrointestinal and 53 pancreatic endocrine tumors.

In four cases, the frozen material was represented by liver or lymph-node metastatic tissue, in the absence of frozen tissue from the primary tumor. In seven additional cases, both primary tumors and the corresponding metastatic deposits were analyzed in parallel using IHC. Finally, in four cases, the pre-operative cytological or biopsy material was also retrieved and studied in parallel with the corresponding surgical sample.

Four carcinoids of the gastric body were associated with chronic atrophic gastritis and neuroendocrine cell hyperplasia [2]. In these cases, also peritumoral normal and hyperplastic neuroendocrine cells were investigated for SSTR expression.

All cases were reviewed on conventional hematoxylin and eosin-stained slides and on slides immunostained for endocrine markers, and classified following the recent World Health Organization (WHO) classification of endocrine tumors [46]. Clinicopathological data including sex, age, tumor location and size, hormonal status and lymph-node or distant metastases were obtained for all patients. Follow-up data were available for 54 patients.

Normal pancreas

Although this study was not specifically designed to study SSTR distribution in normal pancreatic islets, peritumoral pancreatic tissue, when present, as well as two control cases of non-neoplastic pancreatic tissue, were also investigated to compare the immunohistochemical reactivity of the currently employed SSTR antibodies with that reported in the literature.

RT-PCR for SSTRs 1-5

Total RNA was extracted from a frozen tumor fragment adjacent to that used for conventional histopathological diagnosis using the Tri Reagent (Molecular Research Center Inc., Cincinnati, Ohio.) extraction kit following manufacturer's recommendations. To avoid contamination with normal peritumoral pancreatic or lymphatic tissue, frozen sections were rapidly stained with conventional hematoxylin and eosin, and the normal tissue possibly present was removed with a surgical blade. The concentration of RNA was estimated by means of spectrophotometry, and RNA degradation assessed using 1% agarose gel electrophoresis. Total RNA (1 µg) was first digested with 10 U RNase-free DNase (Boehringer, Mannheim, Germany) in a 10-µl solution containing 2 mM MgCl₂ to avoid DNA contamination. The solution was kept at room temperature (RT) for 10 min, then heated for 5 min at 70°C to inactivate DNase; 40 pM of oligodeoxythymidine primer (oligo dT₁₆) was added and the solution heated at 70°C for 10 min, then chilled on ice to allow primer hybridization. The final solution was reverse transcribed using 100 U superscript reverse transcriptase (Gibco BRL, Gaithersburg, Md.), and complementary DNA (cDNA) was generated in a 50-µl final reaction volume containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 1 mM dNTPs and 20 U RNasin (Promega, Madison, Wis.). The solution was incubated at 37°C for 90 min, then the enzymes inactivated by heating at 70°C for 10 min. Omission of RT enzyme served as a negative control for further PCR amplification reactions. RNA integrity was determined by performing a PCR reaction for β₂-microglobulin [7]. Only cases positive for β₂-microglobulin mRNA amplification were considered informative for SSTR expression analysis.

PCR experiments were carried out in a 10-µl final reaction volume containing 1 µl cDNA template, 1 pM sense and antisense oligonucleotide primers, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 200 mM deoxynucleotide triphosphates (dNTPs), 1 mM MgCl₂ and 0.5 U Taq polymerase (GeneAmp, Perkin Elmer, Roche, Branchburg, N.J.). The primers used for RT-PCR and the corresponding specific annealing temperatures are reported in Table 1. β₂-microglobulin and SSTR2 PCR reactions were performed under the following conditions: 35 cycles, each cycle consisting of denaturation at 94°C for 1 min, annealing at specific temperatures for 1 min and extension at 72°C for 2 min. PCR conditions for SSTR3 (35 cycles) and SSTR5 (40 cycles) were as follows: denaturation at 94°C for 45 s, annealing at specific temperatures for 1 min and extension at 72°C for 1 min. SSTR1 and SSTR4 PCR reactions were performed under the following conditions: 40 cycles, each cycle consisting of denaturation at 94°C for 1 min, annealing at specific temperatures for 30 s and extension at 72°C for 75 s. PCR products were visualized under UV light in 1% agarose gel containing 0.5 µg/ml ethidium bromide. Omission of cDNA in

Table 1 Sequences of primers used for reverse transcriptase-polymerase chain reaction (PCR)

Primer	Position	GenBank accession no.	Size of PCR product (bp)	Annealing temperature (°C)	References
β2-Microglobulin					
Sense: 5'acccccactgaaaagatga3'	284–303	XM_007650	114	55	[7]
Antisense: 5'atcttcaaacctccatgatg3'	378–397				
SSTR1					
Sense: 5'gctacgtgctcatcattgcta3'	725–745	NM_001049	401	60	[20]
Antisense: 5'ggactccaggttctcaggttg3'	1105–1125				
SSTR2					
Sense: 5'cagtcagcagcagcagcga3'	402–421	NM_001050	283	61	[50]
Antisense: 5'gcagaacagatgatggtga3'	665–684				
SSTR3					
Sense: 5'tcatctgcctctgctacctg3'	662–681	NM_001051	221	57	[20]
Antisense: 5'gagcccaagaagcagcagct3'	863–882				
SSTR4					
Sense: 5'atcttcgcagacaccagacc3'	547–576	NM_001052	321	60	[20]
Antisense: 5'atcaaggctggtcacgaca3'	848–867				
SSTR5					
Sense: 5'gccggcctctactcttcctg3'	847–866	NM_001053	153	60	[20]
Antisense: 5'ccgtggcgtcagcgtccttg3'	980–999				

the PCR mixture served as negative control. The RNA extracted from an H716 neuroendocrine colon carcinoma cell line known to express SSTR served as a positive control for SSTR amplification. To confirm the specificity of SSTR PCR reactions, PCR products were agarose-gel purified (Concert Gel Extraction Systems, Gibco BRL, Life Technologies, Milan, Italy) and ethanol precipitated, then subjected to nucleic acid direct sequencing using a fluorescence-based sequencing kit (Applied Biosystems, Weiterstadt, Germany); electrophoresis and data analysis were carried out in an automated DNA sequencer ABI PRISM 377 (Applied Biosystems), following manufacturer's recommendations. The RT-PCR was not quantitative. Nevertheless, we used the same controls as those used in a former study on SSTR2 mRNA expression in neuroblastomas using quantitative RT-PCR [44]. Control tissues found in the previous tests to express a number of SSTR2 mRNA molecules lower than 10^5 were negative in the present RT-PCR analysis, while the presence of a strong amplification band corresponded to a number of SSTR2 mRNA molecules larger than 10^7 . No undefined bands were found in the present cases, and the tumors were recorded as positive or negative based on the presence or absence of a specific band, irrespective of its intensity.

Antibodies-SSTR2

A polyclonal antibody anti-SSTR2 was produced and characterized by one of us (MS) [40, 41]. This antibody (coded K230) was raised in sheep and was specific for a sequence of 20 amino acids of the C-terminal portion of the SSTR2A (KSRLNETTETQRTLLNEDLQ, positions 347-366). Another antibody to SSTR2A, purchased from Biotrend/Gransch Laboratories (Schwabhausen, Germany) was also used and found to have an overlapping reactivity pattern with the former SSTR2 antiserum. The antibody to SSTR3 was purchased from Biotrend/Gransch Laboratories. The characterization of these latter antibodies was reported by Schultz et al [43]. The serum anti-SSTR5 (coded MK86) was produced in the laboratory of one of us (MS) against the human SSTR5 receptors (see below).

Production and characterization of anti-SSTR5 antibodies

The following peptide sequence was synthesized (amino acid single letter code): CGAKDADATEPRPDR (corresponding to a car-

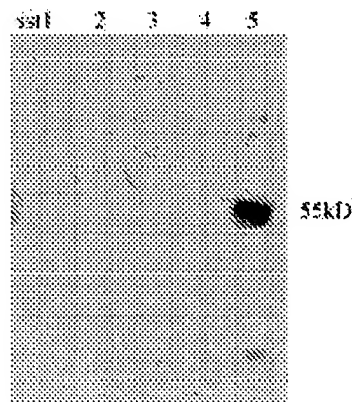


Fig. 1 Western blot of anti-human somatostatin receptor (SSTR)5 antibody (MK86) reactivity in transfected Chinese hamster ovary (CHO) cells expressing SSTR1-5 (columns 1-5). A specific band is present in SSTR5-transfected cells only

boxy-terminal fragment of human SSTR5, positions 325-339). The peptide was coupled to keyhole limpet hemocyanin (KLH; Sigma, Poole, UK) and injected into one rabbit (MK86). Test bleeds were taken after each booster injection, and a terminal bleed was carried out. Serum was obtained from this bleed and used in IHC and Western blotting. The specificity of the novel SSTR5 antiserum was confirmed by means of Western-blot experiments on membranes obtained from Chinese hamster ovary (CHO) cells expressing the recombinant human SSTR types 1-5, carried out as described elsewhere [40]. Application of anti-human SSTR5 antiserum yielded a signal in CHO cells expressing the respective SSTR5 only (Fig. 1). In these experiments, a broad band with smeared appearance was observed with a molecular weight centered at approximately 55 kDa. No cross-reactivity was observed to any other cell line expressing recombinant SSTRs different from type 5 (Fig. 1). Pre-adsorption controls were carried out by pre-incubating the antibodies with 100 μ M of the antigenic

Table 2 Summary of clinico-pathological features of 81 gastroenteropancreatic endocrine tumors

	Gastrointestinal	Pancreatic
Total no. of cases	28	53
Diagnosis (WHO 2000):		
Well-differentiated endocrine tumor	6	37
Well-differentiated endocrine carcinoma	19	14
Poorly differentiated endocrine carcinoma	3	2
Hormonal status (no. of cases)	Gastrin (3)	Insulin (25), glucagon (5), somatostatin (4), gastrin (5), pancreatic polypeptide (1), vasoactive intestinal peptide (1), non functioning (14)
Sex male/female	14/14	25/28
Age (median, years)	60	56
Tumor size (mean, cm)	2.8	4.0
Location	Stomach (6), ileum (13), appendix(4), colon/rectum (5)	Head (22), Body-tail (31)
Metastases (regional lymph node or liver)	16/28 (57.1%)	17/53 (32%)
Studied using reverse-transcriptase polymerase chain reaction and immunohistochemistry	13	33
Studied using immunohistochemistry only (Table 3)	15	20

peptide for 16–18 h prior to immune reactions and resulted in the abolition of the signal in either Western blotting or immunohistochemical studies, using human pituitary gland formalin-fixed sections as positive controls (data not shown).

Immunohistochemistry

Sections (4- to 5- μ m-thick) from buffered formalin-fixed, paraffin-embedded tissue, adjacent to those used for conventional histopathological and immunohistochemical examination, were collected onto poly-L-lysine-coated slides. Primary antibodies to SSTR2A (commercial serum), SSTR3 and SSTR5 were applied to tissue sections with prior antigen retrieval (three 3-min passages in microwave oven at 750 W in 10 mM citrate buffer pH 6.0) for 30 min at RT, at the dilutions of 1/20000, 1/40000 and 1/35000, respectively. The other anti-SSTR2A antiserum (K230) was applied overnight at 4°C at a dilution of 1/2000 with no prior antigen retrieval. The immune reactions were then revealed with the immunoperoxidase technique using the streptavidin-peroxidase kit (Dako, Glostrup, Denmark) and a further amplification with biotinylated-tyranide (GenPoint, Dako, diluted 1/5) with the subsequent washing step performed in hot PBS (at 95°C) to reduce the background and increase the sensitivity, as described by Volante et al. [48]. Diaminobenzidine was used as chromogen, followed by a weak nuclear counterstain (or no counterstain in selected cases). Control stainings for all antibodies included immunoperoxidase reaction on serial sections using pre-absorbed serum or buffer instead of the primary antibody. All pancreatic endocrine tumors were immunostained in parallel sections for insulin (polyclonal, diluted 1/400, Dako) glucagon (polyclonal, diluted 1/5, BioGenex, San Ramon, Calif.) somatostatin (polyclonal, diluted 1/3000, Serotec, Oxford, UK) and gastrin (polyclonal, diluted 1/50, BioGenex).

Double immunostaining for each SSTR subtype and individual pancreatic hormones was performed in the two control pancreatic tissues and in selected tumors with peritumoral islets available for co-localization studies. A sequential immunohistochemical procedure under the same conditions described above was performed, finally revealing the receptor proteins with green fluorescence and the pancreatic hormones of alpha, beta or delta cells with red fluorescence. For the purpose of co-localization analysis of receptor types and single pancreatic hormones, 250 cells (for each cell type) were counted in nearby islets and the relative percentage of cells co-localizing either SSTR type was recorded.

In situ hybridization

Selected tumors (12 cases, including 7 tumors with discrepant RT-PCR and IHC findings) were also analyzed for SSTR2 mRNA expression by means of a non-radioactive, tyramide deposition-based ISH technique. The detailed procedure was described elsewhere [24] and employed a 48-bp oligonucleotide corresponding to positions 91–139 of the SSTR2 gene [50]. Appropriate negative controls in all the cases investigated included hybridization of parallel sections using an unrelated probe specific for Epstein-Barr virus mRNA. The specificity of the hybridization signal was also confirmed in some cases using a sense probe, omitting the specific probe or pre-treating the sections with RNase.

Clinico-pathological correlation

The results of SSTR analysis were compared with clinico-pathological data, including age, sex, tumor location and size, recurrence or metastases, tumor grade, and hormonal status. Statistical evaluation was performed by means of the ANOVA and chi-square (Yates corrected) tests. The data were also related to the outcome in those patients in which this information was available. This study was performed in accordance with the ethics standards laid down in the 1964 Declaration of Helsinki.

Results

Clinico-pathological data

The relevant clinico-pathological data of 53 pancreatic and 28 gastrointestinal endocrine tumors are summarized in Table 2. The majority of cases were well-differentiated endocrine tumors (carcinoids) or well-differentiated endocrine carcinomas. Five cases of poorly differentiated endocrine carcinomas (with small or large cell features) were also included. The male to female ratio was approximately 1 in both groups of pancreatic and intestinal tumors, with the patients' median age being 56 years and 60 years in the two groups, respectively. The pancreatic and gastrointestinal tumors measured 0.4–20 cm

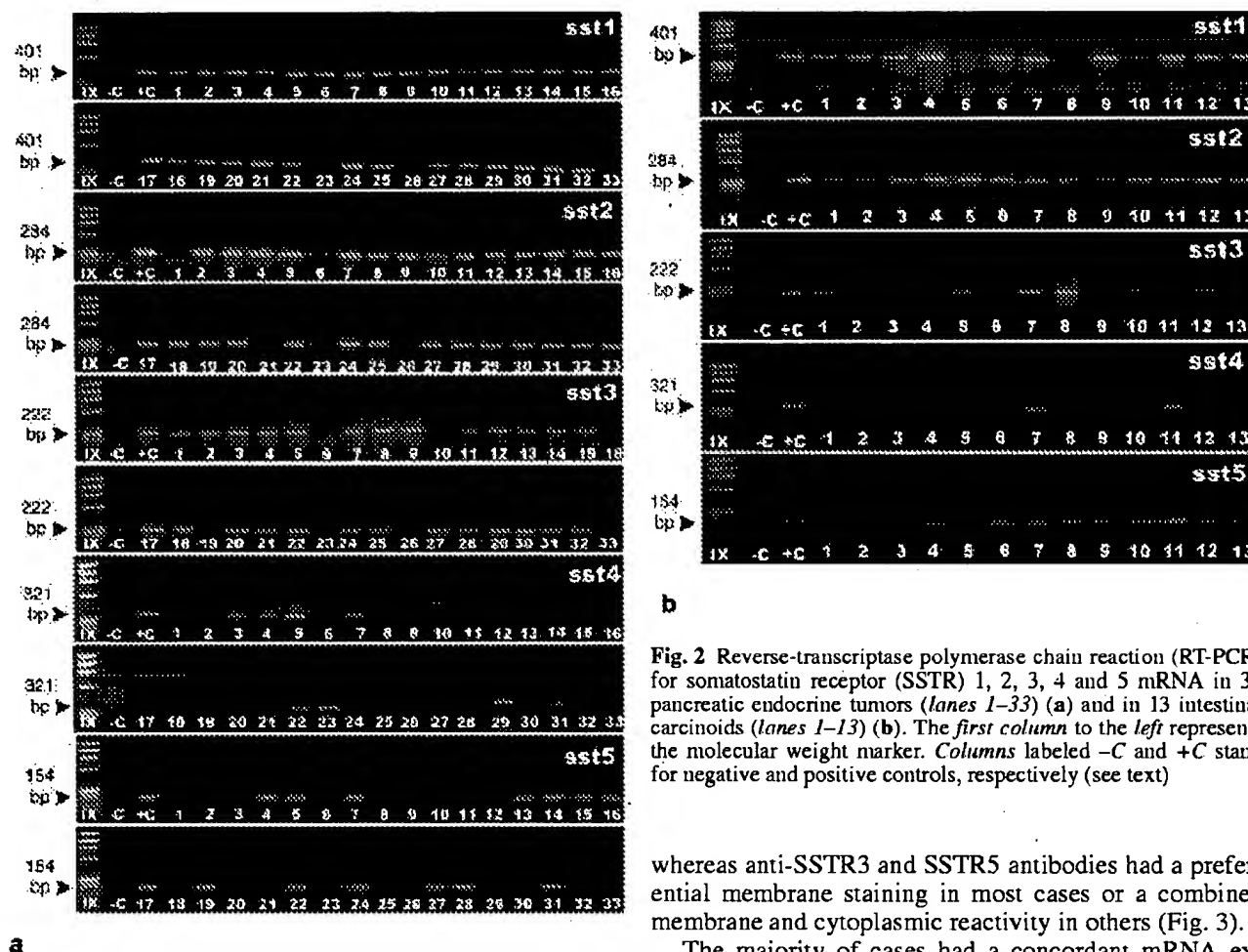


Fig. 2 Reverse-transcriptase polymerase chain reaction (RT-PCR) for somatostatin receptor (SSTR) 1, 2, 3, 4 and 5 mRNA in 33 pancreatic endocrine tumors (lanes 1–33) (a) and in 13 intestinal carcinoids (lanes 1–13) (b). The first column to the left represents the molecular weight marker. Columns labeled –C and +C stand for negative and positive controls, respectively (see text)

(mean 4.0 cm) and 0.3–11 cm (mean 2.8 cm), respectively. Lymph-node and/or liver metastases were observed in 57.1% of intestinal tumors and in one-third (33.9%) of pancreatic tumors.

SSTR in pancreatic endocrine tumors

RT-PCR analysis, performed in 33 cases, showed mRNA amplification of SSTR1 in 30 of 33 (90.1%), of SSTR2 in 28 of 33 (84.8%), of SSTR3 in 26 of 33 (78.8%), of SSTR4 in 8 of 33 (24.2%) and of SSTR5 in 14 of 33 (42.4%) cases (Fig. 2a, Table 3). In all cases, the preservation of total mRNA was ensured by detectable amplification of β 2-microglobulin mRNA. IHC was performed using anti-SSTR 2, 3 and 5 antibodies. The immunohistochemical reactivity was markedly heterogeneous, being in some cases focally positive (in less than 25% of tumor cells, scored one plus, +; Table 3) and in others diffusely immunostained in over 50% of tumor cells (score +++). Some cases had an intermediate reactivity with between 25% and 50% of tumor cells positive (score ++). A cell membrane distribution of the immunoreactivity was present with anti-SSTR2 antibodies,

whereas anti-SSTR3 and SSTR5 antibodies had a preferential membrane staining in most cases or a combined membrane and cytoplasmic reactivity in others (Fig. 3).

The majority of cases had a concordant mRNA expression, as detected by RT-PCR analysis, and protein expression, as shown by IHC. Discrepant cases included two tumors that failed to stain with anti-SSTR2 antibodies, despite a positive RT-PCR. In addition, a focal immunoreactivity for SSTRs 2, 3 and 5 was observed in one, one and four cases, respectively, in the absence of the corresponding mRNA amplification. This is likely due to tumor heterogeneity, since most cases had a focal SSTR immunoreactivity and, although care was taken to freeze tumor fragments adjacent to those used for conventional histopathological analysis, it cannot be excluded that the tumor cell populations of the two samples had a divergent differentiation pattern.

IHC was performed with anti-SSTR2, 3 and 5 antibodies in 22 additional tumors from 20 patients, in which frozen tissues were not available. Two patients (case nos. 35 and 39 – Table 3) had two separate tumor nodules producing a different hormone. For the purpose of correlative analysis between SSTR expression and hormone production, they were kept separate. SSTR2 immunoreactivity was found in 15 of 22 tumors (68.2%), SSTR3 in 8 of 22 (36.4%) and SSTR5 in 14 of 22 (63.6%).

Taking together all data on SSTR expression, it appeared that the majority of pancreatic endocrine tumors were expressing multiple SSTR subtypes, being at least

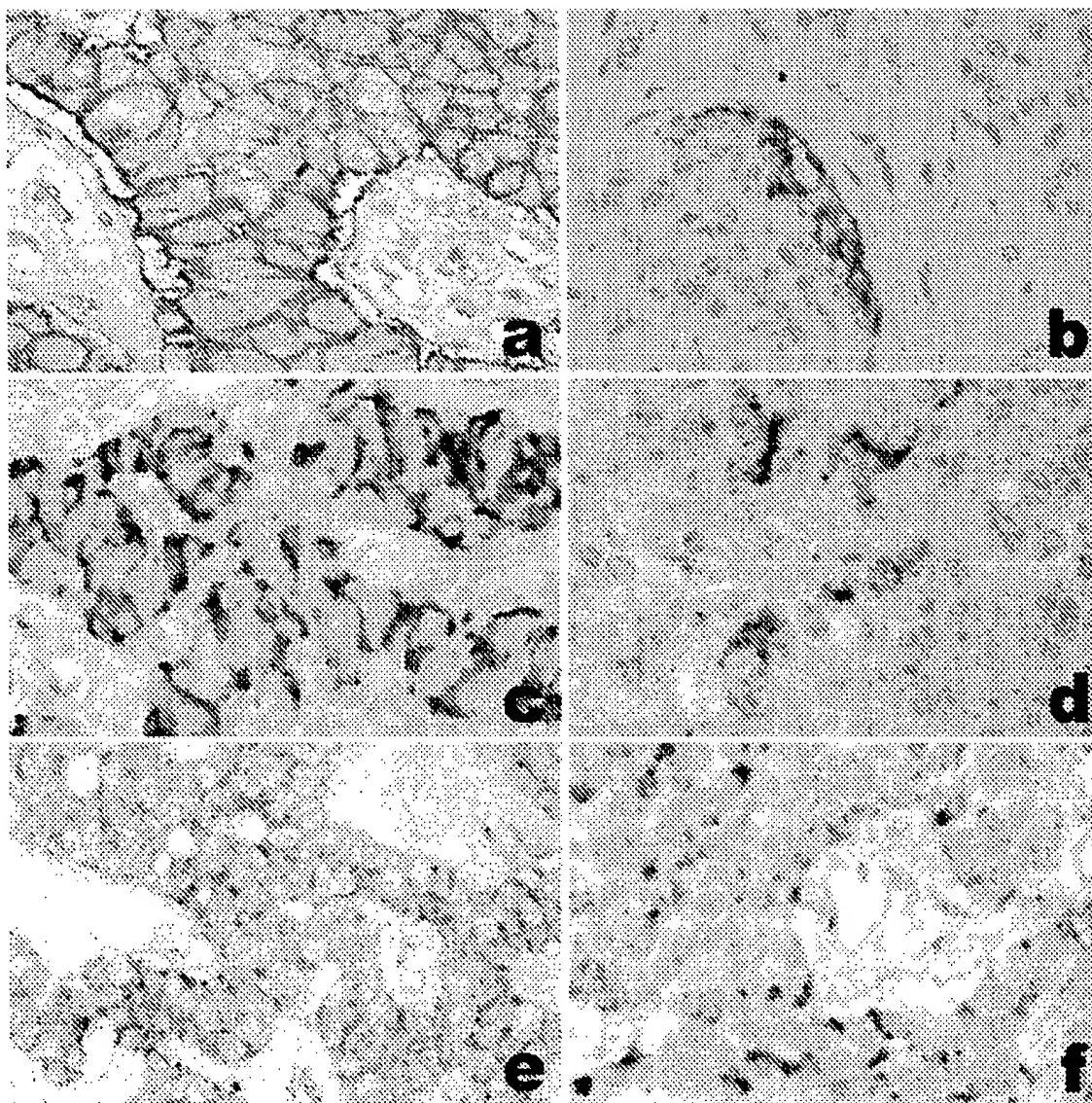


Fig. 3 **a** Case 32. Pancreatic gastrinoma diffusely positive (score +++) for somatostatin receptor (SSTR)2A at the cell membrane level. **b** Case 31. Pancreatic insulinoma showing a focal (score +) positivity for SSTR2A with predominant membranous pattern. **c** Case 31. Strong membrane staining for SSTR3 in most tumor cells (score +++). **d** Case 3. A weak cytoplasmic staining in single tumor cells with membrane reinforcement is present in a pancreatic insulinoma (score +). **e** Case 37. Diffuse cytoplasmic staining with membrane reinforcement in a pancreatic somatostatinoma using the novel MK86 antibody to SSTR5 (score +++). **f** Case 15. Heterogeneous SSTR5 staining in a non-functioning pancreatic tumor with both cytoplasmic and membranous pattern (score ++). (**a-f** $\times 200$). **g-i** Case 8. The pre-operative fine needle aspirate of this insulinoma metastatic to the liver shows a tumor cell cluster in

the cell block (**g**) positive for chromogranin A (**h**) and for SSTR2A, which is localized at the cell membrane of most tumor cells (**i**) (**g-i** $\times 400$). **l-n** Case 14. Non-functioning pancreatic tumor. The same tumor area expresses SSTR2 at protein level (positive SSTR2A immunoperoxidase) (**l**) and also at mRNA level (positive ISH reaction) (**m**). No staining is present hybridizing a parallel section with an unrelated probe (**n**) (**l-n** $\times 400$). **o** Expression of SSTR in normal pancreatic islets. Relative percentages of SSTR2, SSTR3 and SSTR5 co-localization in normal pancreatic islet cells (*left*). An example of double immunofluorescence staining of somatostatin in delta cells (*red*) and SSTR5 (*green*) is shown to the *right*. All delta cells are positive for SSTR5 (*yellow color*), showing either cell membrane (*arrows*) or cytoplasmic (*arrowheads*) staining ($\times 400$)

three SSTR types present in 75% of cases and being only four of 55 tumors completely negative.

The SSTR immunoreactivity pattern was overlapping in lymph-node or liver metastases studied in parallel with the corresponding primary tumor. In addition, the pre-operative biopsy material (fine needle aspirate of a lymph-node or liver metastasis) of three cases gave over-

lapping results relative to the corresponding surgically resected tumor (Fig. 3).

In six cases, ISH for SSTR2 mRNA was performed and showed a concordant pattern with respect to RT-PCR data. A diffuse cytoplasmic positivity was present in the positive cases (Fig. 3), but not in parallel sections stained with an unrelated or sense probe.

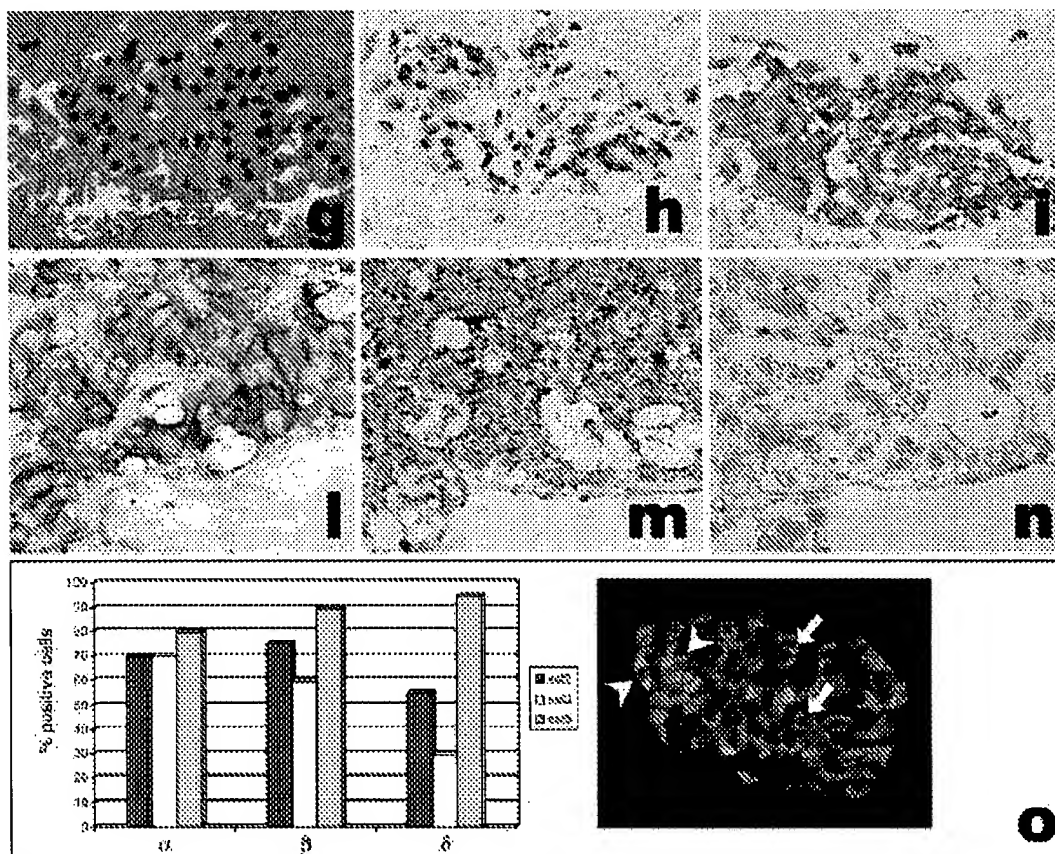


Fig. 3 g-o

When the SSTR expression was evaluated with regard to hormonal activity of the tumors, it turned out that SSTRs were heterogeneously distributed among different functioning or non-functioning endocrine tumors (Fig. 3). Generally, SSTR1 was diffusely expressed in endocrine tumors of the pancreas, as well as SSTR2, which was present in all glucagon (five cases) and gastrin-producing tumors (eight cases, including three located in the duodenum). SSTR3 and SSTR5 were expressed in approximately two-thirds of tumors, with no apparent restriction to a given hormonally active tumor, with the only exception of SSTR5 expressed in all four somatostatin-producing tumors. Finally, SSTR4 was poorly expressed in pancreatic endocrine tumors, being the specific mRNA present in only 8 of 33 tumors (24.2%). Interestingly, all such cases were insulinomas. Non-functioning tumors were not specifically associated with a specific SSTR subtype, irrespective of their size, tumor grade or biological behavior. Two cases, both high grade and metastatic tumors, were negative for all SSTR subtypes.

Normal pancreatic islets

The immunohistochemical expression of SSTRs 2, 3 and 5 was controlled in peritumoral pancreatic islets and in two control cases of normal pancreas, using double

stainings with two-color immunofluorescence. SSTR2 was present in all cell types, co-localizing with 70% of alpha cells, 75% of beta cells and 55% of delta cells. SSTR3 co-localized with 70% of alpha cells, 60% of beta cells and less than one-third of delta cells. SSTR5 was present in 80% of alpha cells, 90% of beta cells, and in virtually all delta cells (95%; Fig. 3).

SSTR in gastrointestinal endocrine tumors

RT-PCR analysis, performed in 13 cases, showed mRNA amplification of SSTR1 in 12 of 13 (92.3%), of SSTR2 in 13 of 13 (100%), of SSTR3 in 7 of 13 (53.8%), of SSTR4 in 2 of 13 (15.4%) and of SSTR5 in 10 of 13 (76.9%) cases (Fig. 2b, Table 4). All the cases analyzed had a detectable level of the housekeeping gene β 2-microglobulin mRNA, thus confirming the total mRNA quality. IHC was performed with anti-SSTR2, 3 and 5 antibodies. The immunohistochemical reactivity was extremely heterogeneous, as for pancreatic tumors. A cell membrane immunoreactivity was present with anti-SSTR2 antibodies, while in SSTR3 and SSTR5 stained cases, cell membrane and cytoplasmic reactivities were combined in some cases (Fig. 4).

The majority of cases had concordant mRNA and protein expression. Discrepant cases had a focal immunoreactivity for SSTR2, 3 and 5 in one, two and one cases, respectively, despite a negative RT-PCR reaction.

468

Table 3 Somatostatin receptor (SSTR) expression in endocrine tumors of the pancreas. *M* male, *F* female, *WDET* well-differentiated endocrine tumor, *WDEC* well-differentiated endocrine carcinoma, *PDEC* poorly differentiated endocrine carcinoma, *PP* pancreatic polypeptide, *VIP* vasoactive intestinal peptide, *LN* lymph node, *NED* no evidence of disease, *DOD* dead of disease, *DOC* dead other causes, *AWD* alive with disease, *IHC* immunohistochemistry, *RT-PCR* reverse-transcriptase polymerase chain reaction, - negative, + positive in <25% of tumor cells, ++ positive in 25-50% of tumor cells, +++ positive in >50% of tumor cells, *nd* not done

No.	Sex/ age	Location/ size (cm)	Diagnosis	Hormonal status	Metastases	Follow-up (years)	SSTR1		SSTR2		SSTR3		SSTR4		SSTR5	
							RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC
1	M/44	Body/1	WDET	Insulin		NED 2	+	-	-	+	+	+	-	-	-	-
2	F/56	Head/1.4	WDET	Insulin		NED 6.5	+	+++	+	++	+	++	+	++	+	++
3	M/58	Body/2	WDET	Insulin		NED 3.5	+	-	+	+	+	+	+	+	+	+
4	F/72	Tail/4	WDET	Insulin		NED 4.5	+	+	+	+++	+	+++	+	+++	+	+++
5	F/39	Body/1.8	WDET	Insulin		NED 4	+	+	+	+	+	+	+	+	+	+
6	F/55	Head/0.9	WDET	Insulin		NED 4	+	-	-	-	-	-	-	-	-	-
7	M/60	Body/1.2	WDET	Insulin		NED 8.5	+	nd	+	nd	+	+	+	+	+	+
8	M/62	Tail/5	WDEC	Insulin	Liver	DOD 1	+	+++	+	+	+	+	+	+	+	+
9	M/50	Tail/4	WDET	PP		NED 6.5	+	+	+	+	+	+	+	+	+	+
10	M/54	Tail/11	WDET	Glucagon		NED 4	+	+	+	+	+	+	+	+	+	+
11	F/63	Tail/7	WDET	VIP	LN	DOD 2	+	+	+	+	+	+	+	+	+	+
12	F/69	Tail/6	WDEC	Insulin		NED 11	+	+	+	+	+	+	+	+	+	+
13	M/44	Head/5	WDET	Non-functioning		DOD 0.1	+	+	+	+	+	+	+	+	+	+
14	F/59	Body-tail/4	WDEC	Non-functioning	LN	NED 4	+	+++	+	+	+	+	+	+	+	+
15	M/49	Head/6	WDET	Non-functioning	Liver	DOC 4.5	+	+	+	+	+	+	+	+	+	+
16	F/54	Head/3.5	WDET	Non-functioning	LN	Lost	+	+	+	+	+	+	+	+	+	+
17	F/44	Body/3	WDEC	Non-functioning	LN	DOD 2	+	+	+	+	+	+	+	+	+	+
18	F/70	Head/3.5	WDET	Non-functioning		NED 1	+	+	+	+	+	+	+	+	+	+
19	F/64	Head/2.3	WDEC	Non-functioning	LN	DOD 1.2	+	+	+	+	+	+	+	+	+	+
20	M/55	Head/2	WDEC	Non-functioning	Liver	NED	+	+	+	+	+	+	+	+	+	+
21	M/56	Body-tail/3	WDET	Insulin		Lost	+	+	+	+	+	+	+	+	+	+
22	F/59	Head/1.5	WDET	Insulin		NED 3	+	+++	+	+	+	+	+	+	+	+
23	M/52	Body/1	WDET	Insulin		NED 11	-	+	+	+	+	+	+	+	+	+
24	F/26	Tail/2	WDET	Insulin		NED 5.5	+	+	+	+	+	+	+	+	+	+
25	M/23	Head/1	WDET	Insulin		Lost	+	-	+	+	+	+	+	+	+	+
26	F/29	Tail/0.5	WDET	Insulin		NED 0.6	-	-	-	-	-	-	-	-	-	-
27	F/56	Body/1	WDET	Insulin		NED 11	+	+	+	+	+	+	+	+	+	+
28	F/61	Tail/1	WDET	Insulin		Lost	+	+	+	+	+	+	+	+	+	+
29	F/72	Head/1.5	WDET	Insulin		Lost	+	+	+	+	+	+	+	+	+	+
30	F/76	Head/2	WDET	Insulin		Lost	+	+	+	+	+	+	+	+	+	+
31	F/59	Body/2.5	WDET	Non-functioning		Lost	+	+	+	+	+	+	+	+	+	+
32	M/28	Body-tail/7.5	WDEC	Gastrin	Liver	AWD 0.5	+	+++	+	+	+	+	+	+	+	+
33	M/76	Head/11	PDEC	Non-functioning	LN	DOD 0.5	-	-	+	-	-	-	-	-	-	-
34	M/62	Tail/1.2	WDET	Insulin		Lost	nd	-	+	-	nd	-	nd	+	+	+
35	F/27	Body-tail/2.5	WDET	Insulin/glucacon		Lost	nd	-	+	-	nd	-	nd	+	+	+
36	F/61	Body/1.5	WDET	Insulin		Lost	nd	-	+	-	nd	-	nd	+	+	+
37	F/55	Head/2	WDET	Somatostatin		Lost	nd	+	+	+	+	+	+	+	+	+
38	M/48	Tail/10	WDET	Somatostatin		Lost	nd	+	+	+	+	+	+	+	+	+
39	M/46	Body-tail/7	WDET	Insulin/somatostatin		DOC 8.5	nd	++/-	+	-	nd	-	nd	+	+	+

Table 3 (continued)

No.	Sex/ age	Location/ size (cm)	Diagnosis	Hormonal status	Metastases	Follow-up (years)	SSTR1		SSTR2		SSTR3		SSTR4		SSTR5	
							RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC
40	F/54	Head/2.5	WDEC	Gastrin	LN	Lost	nd	+	nd	-	nd	-	nd	+	nd	
41	F/59	Head/3.5	WDET	Gastrin		Lost	nd	++	nd	-	nd	-	nd	++	nd	
42	M/58	Head/4	WDEC	Non-functioning	LN	DOD 1	nd	-	nd	-	nd	-	nd	-	nd	
43	M/67	Head/0.6	WDET	Glucagon	LN	DOD 1	nd	+	nd	-	nd	-	nd	+	nd	
44	F/67	Body/0.4	WDET	Glucagon		DOD 0.4	nd	+	nd	+	nd	+	nd	+	nd	
45	M/54	Head/3	WDET	Insulin		NED 9	nd	-	nd	-	nd	-	nd	-	nd	
46	M/55	Head/2.8	WDEC	Gastrin	LN	Lost	nd	+	nd	+	nd	+	nd	+	nd	
47	F/57	Tail/1.3	WDET	Insulin		NED 7.5	nd	+	nd	-	nd	-	nd	-	nd	
48	F/39	Head/7	WDEC	Gastrin	LN	DOD 14	nd	++	nd	+	nd	+	nd	++	nd	
49	M/54	Body/6	WDEC	Glucagon	Liver, lung	DOD 1	nd	+	nd	-	nd	-	nd	+	nd	
50	M/46	Tail-spleen/12	WDEC	Somatostatin		NED 2	nd	-	nd	-	nd	+	nd	+	nd	
51	M/20	Body-tail/20	PDEC	Non-functioning	LN	DOD 1	nd	-	nd	-	nd	-	nd	-	nd	
52	F/59	Body-tail/3.5	WDEC	Non-functioning		NED 2	nd	++	nd	+	nd	+	nd	+	nd	
53	M/57	Head/2	WDET	Non-functioning		NED 1	nd	++	nd	++	nd	++	nd	++	nd	

IHC alone was performed with anti-SSTR2, 3 and 5 antibodies in 15 additional tumors, 8 of which had lymph-node and/or liver metastases. SSTR2 immunoreactivity was found in 12 of 15 tumors (80%), SSTR3 in 9 of 15 (60%) and SSTR5 in 12 of 15 (80%). In four cases of gastric body carcinoid associated with chronic atrophic gastritis and hyperplasia of neuroendocrine ECL cells, small hyperplastic nests as well as single endocrine cells scattered in the gastric glands were also positive for SSTR2 immunostaining (Fig. 4).

Taking together all data on SSTR expression, it appeared that the majority of gastrointestinal endocrine tumors were expressing multiple SSTR subtypes, with at least three SSTR types present in 90% of cases (with one tumor expressing all five types and three others having four of five SSTR types expressed) and with only one case completely negative (a poorly differentiated endocrine carcinoma).

As for the pancreatic tumors, the SSTR immunoreactivity pattern was overlapping in lymph-node or liver metastases relative to that of the primary tumor, as well as in pre-operative biopsy material (fine needle aspirate of a liver metastasis; Fig. 4) and in the corresponding surgically resected tumor.

In six cases, ISH for SSTR2 mRNA was performed and showed a concordant pattern with respect to RT-PCR data. A diffuse cytoplasmic staining was present in the positive cases, but not in parallel sections stained with an unrelated or sense probe (Fig. 4).

Clinico-pathological correlation

No statistically significant correlation was observed between SSTR expression (any type) and either sex, age, tumor size, tumor location or biological behavior. With regard to tumor grade, the rare poorly differentiated endocrine carcinomas (five cases) appeared to express SSTR to a lower extent, although the figures are too small to allow a definite conclusion on this point. Comparing the extent of SSTR expression with the clinical outcome in the 54 patients in which this information was

Fig 4 a-c Case 8. Duodenal gastrinoma metastatic to a lymph node. Somatostatin receptor (SSTR)2 is diffusely expressed as detected by positive immunoreaction (a, score ++++) and positive in situ hybridization (ISH) for SSTR2 mRNA (b). No stain is present in a parallel section hybridized with an unrelated probe (c). (a-c $\times 200$). **d** Case 9. Ileal endocrine tumor showing strong membranous SSTR3 immunoreactivity in most tumor cells (score ++++) ($\times 400$). **e** Case 16. Ileal endocrine tumor with a predominant membranous pattern of staining with the SSTR5 antibody (score +++). A normal intestinal gland (top) is negative ($\times 400$). **f-g** Case 28. A gastric carcinoid associated with chronic atrophic gastritis and neuroendocrine ECL-cell hyperplasia (as shown by chromogranin A immunostaining) (f) are both positive for SSTR2 immunoreaction (g, arrows). (f-g $\times 200$); **h-j** Case 6. The pre-operative fine needle aspirate of an ileal carcinoid metastatic to the liver shows tumor cell clusters in the cell block (h) diffusely positive for SSTR2A (i) with the staining localized at the cell membrane (j) (h-i $\times 200$, j $\times 400$).

470

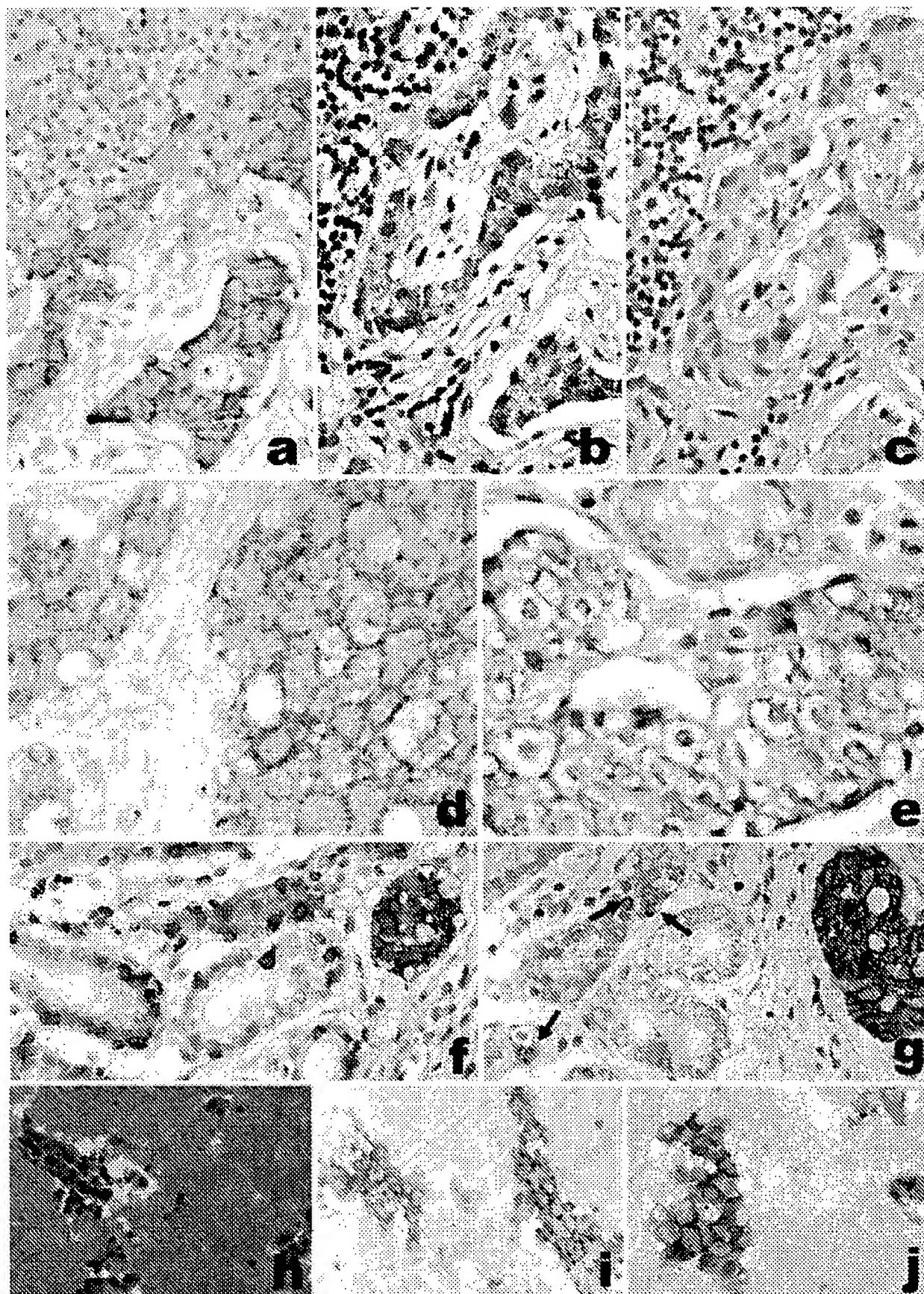


Fig 4 a-c Legend see page 469

Table 4 Somatostatin receptor (SSTR) expression in endocrine tumors of the gastrointestinal tract. *M* male, *F* female, *WDET* well-differentiated endocrine tumor, *WDEC* well-differentiated endocrine carcinoma, *PDEC* poorly differentiated endocrine carcinoma, *LN* lymph node, *NED* no evidence of disease, *DOD* dead of disease, *DOC* dead other causes, *AWD* alive with disease, *IHC* immunohistochemistry, *RT-PCR* reverse-transcriptase polymerase chain reaction, - negative, + positive in <25% of tumor cells, ++ positive in 25-50% of tumor cells, +++ positive in >50% of tumor cells, *nd* not done

No.	Sex/ age	Location/ size (cm)	Diagnosis	Hormonal status	Metastases	Follow-up (years)	SSTR1		SSTR2		SSTR3		SSTR4		SSTR5	
							RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC	RT-PCR	IHC
1	F/52	Appendix/1.5	WDEC		Peritoneum	DOC 0.1	+	+++	+	+	+	+++	-	-	-	-
2	F/68	Appendix/1.5	PDEC			DOC 6	+	-	+	+	+	-	-	-	-	-
3	M/40	Stomach/7	WDEC			NED 7.5	+	+++	+	-	-	+++	-	-	+	+
4	M/69	Ileum/2	WDEC		Liver		+	-	+	+	-	+	-	-	+	+
5	M/66	Ileum/4	WDEC		Peritoneum	NED 5.5	+	+	+	+	+	+	-	-	+	+
6	F/34	Ileum/2	WDEC		Liver	NED 1	+	+	+	+	-	+	-	-	+	+
7	M/62	Ileum/1	WDEC			NED 5	+	+++	+	+	+	+	+	+	+	+
8	M/50	Duodenum/2	WDEC	gastrin	LN		-	+++	+	+	+	+	-	-	+	+
9	M/54	Ileum/1.5	WDEC		Liver, LN	NED 1	+	+++	+	+	+	+	-	-	+	+
10	F/69	Ileum/1.5	WDEC			NED 3.5	+	+	+	+	+	+	-	-	+	+
11	F/55	Duodenum/1.5	WDEC	gastrin		NED 6	+	+	+	+	+	+	+	+	+	+
12	M/52	Colon/3	WDEC		LN		+	nd	+	+	+	+	-	-	+	+
13	F/47	Rectum/1	WDEC		Liver		+	nd	+	+	+	+	-	-	+	+
14	M/31	Duodenum/0.8	WDEC	gastrin	LN	NED 7.5	nd	+	nd	+	+	+	nd	nd	+	+
15	F/29	Rectum/1.1	PDEC		LN		nd	-	nd	-	nd	-	nd	nd	nd	nd
16	F/69	Ileum/5	WDEC		LN	NED 6	nd	+	nd	+	+	+	nd	nd	nd	nd
17	M/74	Colon/3.5	WDEC		LN		nd	+	nd	+	+	+	nd	nd	nd	nd
18	M/66	Stomach/0.2	WDET				nd	+	nd	+	+	+	nd	nd	nd	nd
19	F/21	Appendix/1	WDET				nd	+	nd	+	+	+	nd	nd	nd	nd
20	F/68	Appendix/0.6	WDEC			NED 3	nd	+	nd	+	+	+	nd	nd	nd	nd
21	M/71	Ileum/2.5	WDEC				nd	+++	+	+	+	+	nd	nd	nd	nd
22	M/60	Ileum/0.8	WDEC		Liver		nd	+++	+	+	+	+	nd	nd	nd	nd
23	M/60	Duodenum/10	WDEC		LN		nd	+	nd	+	+	+	nd	nd	nd	nd
24	F/67	Stomach/1.5	WDET			NED 2	nd	-	nd	-	nd	-	nd	nd	nd	nd
25	F/64	Stomach/3.8	WDEC		LN	NED 0.8	nd	+	nd	+	+	+	nd	nd	nd	nd
26	M/66	Rectum/7	PDEC		LN	DOD 2	nd	+	nd	+	+	+	nd	nd	nd	nd
27	F/57	Stomach/0.3	WDET				nd	+++	+	+	+	+	nd	nd	nd	nd
28	F/50	Stomach/0.5	WDET			NED 5	nd	+	+	+	+	+	nd	nd	nd	nd

available, it appeared that no specific association was present between SSSTR content and biological behavior. Finally, comparing pancreatic and gastrointestinal endocrine tumors, it can be noticed that the percentage of SSSTR subtype expression in the two groups is relatively similar, with SSSTRs 1 and 2 present in the vast majority of GEP tumors, SSSTRs 3 and 5 expressed by approximately two-thirds of cases and SSSTR 4 only occasionally represented in these tumors (mostly insulinomas).

Discussion

IHC using specific antibodies is a reliable tool to reveal SSSTR

In 46 GEP endocrine tumors, SSSTR 1-5 expression was found in a consistent percentage of cases with the exception of SSSTR4, which was poorly represented (20%). These findings are the result of RT-PCR and immunohistochemical analyses. The literature data, mostly obtained from small case series (with few exceptions) [12, 49] are largely heterogeneous both in terms of SSSTR expression results and methods employed to demonstrate them. On the contrary, we analyzed SSSTR types 1-5 in a relatively large series of tumors and compared two sensitive techniques - RT-PCR and IHC with tyramide amplification. With regard to the former, few studies have detected SSSTR types 1-5 mRNA in GEP tumors. While most authors agree on a high (>60%) SSSTR1 and 2 and moderate (50%) SSSTR5 expression, conflicting results were observed for SSSTRs 3 and 4, which were found in nearly half of the cases of endocrine pancreatic tumors by some authors [12, 32], but very poorly represented according to others [49]. IHC, however, is the last entry in the list of techniques used to reveal SSSTR, and takes advantage from the possibility of a topographical SSSTR localization within tumor cells using highly specific antibodies. So far, only SSSTR2A immunoreactivity has been thoroughly characterized in GEP endocrine tumors, with other SSSTR types so far having been detected by mRNA analysis alone.

In the current study, SSSTRs 2, 3 and 5 were studied by means of IHC with specific antibodies and were found to correlate with RT-PCR data in over 90% of cases, supporting the data already published on SSSTR2A immunohistochemical detection [14, 33] and expanding them to SSSTR types 3 and 5. The latter were detected using a commercial antibody to SSSTR3 and a novel antibody to SSSTR5 developed in our laboratory. Both SSSTR3 and SSSTR5 immunostainings showed a preferential cell membrane subcellular distribution, as typically observed for SSSTR2A [24, 33]. In some cases, however, an associated cytoplasmic reactivity was also found, as described for SSSTR2A in tumors of other organs, too [38]. A similar pattern of reactivity for SSSTRs 2, 3 and 5 was also observed in biopsy or fine needle aspirate cytology specimens obtained pre-operatively from metastatic liver nodules or primary pancreatic tumors, relative to the corresponding surgical samples.

This indicates that the SSSTR profile can be investigated using IHC in the course of the diagnostic work-up of a GEP tumor directly on small pre-operative biopsy specimens. A negative result, however, might not definitely exclude the presence of a given SSSTR type, since 20-30% of operated cases had a focal immunoreactivity and the biopsy sample may not contain SSSTR immunoreactive cells. In any case, a positive immunohistochemical SSSTR detection may open new prospects for somatostatin analogue-based radio-imaging and radiotherapy of GEP endocrine tumors.

Interpretation criteria of SSSTR expression need to be defined

Cases showing discrepant results between IHC and RT-PCR accounted for up to 10% of cases in each SSSTR type (namely, four, three and five cases for SSSTRs 2, 3 and 5, respectively). These included two tumors positive by RT-PCR alone that failed to show the corresponding SSSTR immunoreactivity, a finding possibly due to poor antigen preservation after fixation or low amount of receptor protein (below the threshold of detectability of immunohistochemical procedure). The remaining cases were negative by RT-PCR and had a focal reactivity (less than 20%) in tumor cells. This unexpected finding is possibly related to tumor heterogeneity, although care had been taken to freeze a fragment adjacent to that used for conventional histopathological examination. Considering weak or focally positive reactions, it remains to be ascertained whether the immunohistochemical evaluation (and probably also SSSTR evaluation in general, with any method) may require a semi-quantitative procedure or a cut-off value, to better interpret borderline findings. An answer to this point will stem from clinical studies on large series of GEP tumors treated with somatostatin analogues. The available literature data [10, 13, 19, 22, 26, 36] showed that a relatively good correlation exists between SSSTR expression (evaluated by autoradiography or RT-PCR) and in vivo detection of functional receptors by means of octreotide scintigraphy, although discrepant cases were also found [13]. Unfortunately, no cases of the present series underwent clinical evaluation with octreotide.

Correlative SSSTR expression in pancreatic islet cells and related endocrine tumors

Although not specifically designed as a study on normal pancreatic islets, we had the opportunity of mapping the individual SSSTR types in peritumoral pancreatic islet cells, as well as in two normal pancreatic tissues used as controls. This was of special interest because of discrepant data on the immunohistochemical localization of SSSTR in pancreatic endocrine cell types. Kumar et al. [18] found SSSTRs 1 and 2 in alpha and beta cells and SSSTR5 in beta and delta cells, with SSSTRs 3 and 4 poor-

ly represented in the pancreas. The data on SSTR2 are in agreement with the findings of Reubi et al. in human [34] and Mitra et al. [21] and Hunyadi et al. [11] in rat pancreas. Conversely, Portela-Gomes and co-workers [28] observed a wide occurrence of SSTR subtypes in all pancreatic endocrine cells, (range 60–100% of individual cell types). In this study, considering specifically only a predominant cell membrane immunoreactivity, we found SSTR2 expressed in 70%, 75% and 55% of alpha, beta and delta cells, respectively; SSTR3 expressed in two-thirds of alpha and beta cells but in less than 30% of delta cells; and finally SSTR5 co-localized in virtually all delta cells and in 80–90% of alpha and beta cells. These figures are in agreement those of Portela-Gomes et al. [28] and Kumar et al. [18] on SSTR5 in delta cells and Reubi et al. [34] and Strowski et al. [47] on SSTR2 in alpha and beta cells. The results of Kumar et al. [18] on other SSTR types are slightly different, possibly due to the fact that these authors used antibodies to rat SSTRs. In particular, SSTR3 expression was very low in the experience of both Kumar et al. [17] and Kubota et al. [16], while both RT-PCR and IHC were strongly positive in our cases.

With regard to SSTR expression in pancreatic neoplasms, SSTR2 was widely expressed by gastrin- and glucagon-producing tumors (100% of cases), including three duodenal gastrinomas. In addition, all somatostatin-producing tumors were SSTR5 positive, a finding which parallels the profile of normal delta cells, as described above. Incidentally, three of four cases had a combined membrane and cytoplasmic positivity, in agreement with the data of Reubi et al. [38] on SSTR2A immunoreactivity in somatostatinomas. As far as insulin-producing tumors are concerned, SSTR2 was present in 76% of tumors, thus confirming that these tumors are rich in SSTR2, just as normal beta cells are [14, 33, 34, 49]. Irrespective of the hormonal product, most tumors expressed SSTR1 (as demonstrated by RT-PCR data only) and two-thirds of them were rich in SSTR3. This is at variance with Wulbrand's findings [49], who showed that SSTR3 was absent in the pancreatic tumors. This difference is surprising since in Wulbrand's [49] and in the present studies, the same primer sequences were used for RT-PCR. Apart from possible technical details (e.g., tissue preservation and storage), other factors related to individual tumors may account for these discrepancies. Only 20% of cases expressed SSTR4, being all such cases insulinomas. Finally, non-functioning tumors did not appear to be associated with a special SSTR expression profile. The relatively large series of 14 cases here studied showed a heterogeneous pattern of SSTR expression, as also observed in the seven non-functioning pancreatic tumors of Wulbrand's series [49]. Nevertheless, the amount of receptors identified in all but two cases (both aggressive tumors) suggests that somatostatin analogue treatment may be a useful adjunct in these tumors, as also suggested by Bartsch [1].

A separate comment is deserved for the two cases with multiple pancreatic nodules producing different

hormones. Interestingly, IHC (and ISH for SSTR2) could demonstrate a selective expression of SSTR subtypes in different nodules, with SSTRs 2 and 5 confined to the insulin and somatostatin-producing nodules, respectively, in one case, and with SSTRs 2 and 3 present in the glucagon-producing tumor, but not in the insulin-positive nodule, of the second case. The clinical implications of this differential SSTR expression in combined endocrine tumors are not fully understood, but such a receptor profile might better address the therapeutic strategy with subtype-selective somatostatin analogues [6, 36, 37], in the case of inoperable disease or recurrent tumors.

The observed correlation between the SSTR subtype(s) expression in normal islet cells and in their related hormone-producing tumors, was also detected in normal/hyperplastic neuroendocrine ECL cells of the oxyntic gastric mucosa and their related carcinoid tumors.

Possible association of reduced SSTR expression and tumor grade

Taking into account clinical and pathological parameters of GEP endocrine tumors, we were unable to disclose any specific association between SSTR subtype expression and either sex, age, tumor location and size, tumor stage or clinical outcome (in the 54 cases in which this information was obtained). In GEP adenocarcinomas, however, an inverse relationship between SSTR2 expression and tumor stage was found [3]. With regard to tumor-grade, as observed in tumors of other locations (e.g., lung) [24], poorly differentiated endocrine carcinomas had an apparently reduced SSTR content. Three of five cases (60%) of poorly differentiated carcinomas had no or only one SSTR type focally expressed in their tumor cell population (as opposed to 12% of well-differentiated cases with only one or no SSTR type expressed).

Conclusion

- SSTR expression in GEP endocrine tumors is heterogeneous; SSTRs 1 and 2 are the types most commonly detected (over 80%), followed by SSTRs 3 and 5 present in two-thirds of cases and by SSTR4 which is poorly represented in GEP tumors.
- RT-PCR is a highly sensitive technique to reveal SSTRs but needs to be combined with methods that allow confirmation of the receptor protein localization in tumor cells.
- Immunohistochemical detection of SSTR2, and also of SSTRs 3 and 5, is a reliable procedure for the study of formalin-fixed tissues, including biopsy and cytological pre-operative material.
- Hormonally active pancreatic tumors have a preferential SSTR5 expression in "somatostatinomas" and SSTR2 positivity in "gastrinomas" and "glucagonomas". "Insulinomas" express all SSTR types (ranging from 40% to 90% of cases). Non-functioning tumors

474

express all SSTR types (but SSTR4) in 60–90% of cases.

- e. A correlation was observed between SSTR subtype expression in normal islet cells and neuroendocrine gastric cells and in the corresponding endocrine tumors.
- f. SSTR expression is not related to tumor stage or outcome but is related to tumor grade, with high-grade endocrine carcinomas having a reduced SSTR content.

Acknowledgements Work was supported by grants from the Italian Ministry of University and Research (ex 60% to MP) and Ministry of Health (ICS 030.12/RF99.28) and from the Italian Association for Cancer Research (AIRC, Milan).

References

1. Bartsch DK, Schilling T, Ramaswamy A, Gerdes B, Celik I, Wagner HJ, Simon B, Rothmund M (2000) Management of nonfunctioning islet cell carcinomas. *World J Surg* 24:1418–1424
2. Bordi C, Annibale B, Azzoni C, Marignani M, Ferraro G, Antonelli G, D'Adda T, D'Ambra G, Delle Fave G (1997) Endocrine cell growths in atrophic body gastritis. Critical evaluation of a histological classification. *J Pathol* 182:339–346
3. Buscail L, Saint-Laurent N, Chastre E, Vaillant JC, Gespach C, Capella G, Kalthoff H, Lluís F, Vaysse N, Susini C (1996) Loss of sst2 somatostatin receptor gene expression in human pancreatic and colorectal cancer. *Cancer Res* 56:1823–1827
4. Corleto VD, Scopinaro F, Angeletti S, Materia A, Basso N, Poletini E, Annibale B, Schillaci O, D'Ambra G, Marignani M, Gualdi G, Bordi C, Passaro EJ, Delle Fave G (1996) Somatostatin receptor localization of pancreatic endocrine tumors. *World J Surg* 20:241–244
5. Eriksson B, Oberg K (1999) Summing up 15 years of somatostatin analog therapy in neuroendocrine tumors: future outlook. *Ann Oncol* 10:S31–S38
6. Fehmann HC, Wulbrand U, Arnold R (2000) Treatment of endocrine gastroenteropancreatic tumors with somatostatin analogues. *Recent Results Cancer Res* 153:15–22
7. Gussow D, Rein R, Ginjaar I, Hochstenbach F, Seemann G, Kottman A, Ploegh HL (1987) The human beta 2-microglobulin gene. Primary structure and definition of the transcriptional unit. *J Immunol* 139:3132–3138
8. Helboe L, Møller M, Norregaard L, Schiødt M, Stidsen CE (1997) Development of selective antibodies against the human somatostatin receptor subtypes sst1–sst5. *Brain Res Mol Brain Res* 49:82–88
9. Hofland LJ, Breeman WA, Krenning EP, de Jong M, Waaijers M, van Koetsveld PM, Macke HR, Lamberts SW (1999) Internalization of [DOTA degrees,125I-Tyr3] octreotide by somatostatin receptor-positive cells in vitro and in vivo: implications for somatostatin receptor-targeted radio-guided surgery. *Proc Assoc Am Physicians* 111:63–69
10. Hofland LJ, Liu Q, Van Koetsveld PM, Zijderwijk J, Van Der Ham F, De Krijger RR, Schonbrunn A, Lamberts SW (1999) Immunohistochemical detection of somatostatin receptor subtypes sst1 and sst2A in human somatostatin receptor positive tumors. *J Clin Endocrinol Metab* 84:775–780
11. Hunyady B, Hipkin RW, Schonbrunn A, Mezey E (1997) Immunohistochemical localization of somatostatin receptor SST2A in the rat pancreas. *Endocrinology* 138:2632–2635
12. Jais P, Terris B, Ruzsiewski P, LeRomancer M, Reyl-Desmurs F, Vissuzaine C, Cadiot G, Mignon M, Lewin MJ (1997) Somatostatin receptor subtype gene expression in human endocrine gastroentero-pancreatic tumours. *Eur J Clin Invest* 27:639–644
13. Janson ET, Stridsberg M, Gobl A, Westlin JE, Oberg K (1998) Determination of somatostatin receptor subtype 2 in carcinoid tumors by immunohistochemical investigation with somatostatin receptor subtype 2 antibodies. *Cancer Res* 58:2375–2378
14. Kimura N, Pilichowska M, Date F, Kimura I, Schindler M (1999) Immunohistochemical expression of somatostatin type 2A receptor in neuroendocrine tumors. *Clin Cancer Res* 5:3483–3487
15. Krenning EP, Bakker WH, Breeman WA, Koper JW, Kooij PP, Aulsema L, Lameris JS, Reubi JC, Lamberts SW (1989) Localisation of endocrine-related tumours with radioiodinated analogue of somatostatin. *Lancet* 1:242–244
16. Kubota A, Yamada Y, Kagimoto S, Shimatsu A, Imamura M, Tsuda K, Imura H, Seino S, Seino Y (1994) Identification of somatostatin receptor subtypes and an implication for the efficacy of somatostatin analogue SMS 201–995 in treatment of human endocrine tumors. *J Clin Invest* 93:1321–1325
17. Kumar U, Laird D, Srikant CB, Escher E, Patel YC (1997) Expression of the five somatostatin receptor (SSTR1–5) subtypes in rat pituitary somatotrophs: quantitative analysis by double-layer immunofluorescence confocal microscopy. *Endocrinology* 138:4473–4476
18. Kumar U, Sasi R, Suresh S, Patel A, Thangaraju M, Metrakos P, Patel SC, Patel YC (1999) Subtype-selective expression of the five somatostatin receptors (hSSTR1–5) in human pancreatic islet cells: a quantitative double-label immunohistochemical analysis. *Diabetes* 48:77–85
19. Lamberts SW, Hofland LJ, van Koetsveld PM, Reubi JC, Bruining HA, Bakker WH, Krenning EP (1990) Parallel in vivo and in vitro detection of functional somatostatin receptors in human endocrine pancreatic tumors: consequences with regard to diagnosis, localization, and therapy. *J Clin Endocrinol Metab* 71:566–574
20. Mato E, Matias-Guiu X, Chico A, Webb SM, Cabezas R, Bernal L, De Leiva A (1998) Somatostatin and somatostatin receptor subtype gene expression in medullary thyroid carcinoma. *J Clin Endocrinol Metab* 83:2417–2420
21. Mitra SW, Mezey E, Hunyady B, Chamberlain L, Hayes E, Foor F, Wang Y, Schonbrunn A, Schaeffer JM (1999) Colocalization of somatostatin receptor sst5 and insulin in rat pancreatic beta-cells. *Endocrinology* 140:3790–3796
22. Nilsson O, Kolby L, Wangberg B, Wigander A, Billig H, William-Olsson L, Fjalling M, Forsell-Aronsson E, Ahlman H (1998) Comparative studies on the expression of somatostatin receptor subtypes, outcome of octreotide scintigraphy and response to octreotide treatment in patients with carcinoid tumours. *Br J Cancer* 77:632–637
23. Panetta R, Greenwood MT, Warszyńska A, Demchyshyn LL, Day R, Niznik HB, Srikant CB, Patel YC (1994) Molecular cloning, functional characterization, and chromosomal localization of a human somatostatin receptor (somatostatin receptor type 5) with preferential affinity for somatostatin-28. *Mol Pharmacol* 45:417–427
24. Papotti M, Croce S, Macri L, Funaro A, Pecchioni C, Schindler M, Bussolati G (2000) Correlative immunohistochemical and reverse transcriptase polymerase chain reaction analysis of somatostatin receptor type 2 in neuroendocrine tumors of the lung. *Diagn Mol Pathol* 9:47–57
25. Papotti M, Kumar U, Volante M, Pecchioni C, Patel YC (2001) Immunohistochemical detection of somatostatin receptor types 1–5 in medullary carcinoma of the thyroid. *Clin Endocrinol (Oxf)* 54:641–649
26. Papotti M, Croce S, Bellò M, Bongiovanni M, Allia E, Schindler M, Bussolati G (2001) Expression of somatostatin receptor types 2, 3 and 5 in biopsies and surgical specimens of human lung tumours. *Virchows Arch* 439:787–793
27. Patel YC (1997) Molecular pharmacology of somatostatin receptor subtypes. *J Endocrinol Invest* 20:348–367
28. Portela-Gomes GM, Stridsberg M, Grimelius L, Oberg K, Janson ET (2000) Expression of the five different somatostatin receptor subtypes in endocrine cells of the pancreas. *Appl Immunohistochem Molecul Morphol* 8:126–132

29. Reubi JC, Hacki WH, Lamberts SW (1987) Hormone-producing gastrointestinal tumors contain a high density of somatostatin receptors. *J Clin Endocrinol Metab* 65:1127-1134
30. Reubi JC, Maurer R, von Werder K, Torhorst J, Klijn JG, Lamberts SW (1987) Somatostatin receptors in human endocrine tumors. *Cancer Res* 47:551-558
31. Reubi JC, Kvols LK, Waser B, Nagorney DM, Heitz PU, Charboneau JW, Reading CC, Moertel C (1990) Detection of somatostatin receptors in surgical and percutaneous needle biopsy samples of carcinoids and islet cell carcinomas. *Cancer Res* 50:5969-5977
32. Reubi JC, Schaer JC, Waser B, Mengod G (1994) Expression and localization of somatostatin receptor SSTR1, SSTR2, and SSTR3 messenger RNAs in primary human tumors using in situ hybridization. *Cancer Res* 54:3455-3459
33. Reubi JC, Kappeler A, Waser B, Laissue J, Hipkin RW, Schonbrunn A (1998) Immunohistochemical localization of somatostatin receptors sst2A in human tumors. *Am J Pathol* 153:233-245
34. Reubi JC, Kappeler A, Waser B, Schonbrunn A, Laissue J (1998) Immunohistochemical localization of somatostatin receptor sst2A in human pancreatic islets. *J Clin Endocrinol Metab* 83:3746-3749
35. Reubi JC, Laissue JA, Waser B, Steffen DL, Hipkin RW, Schonbrunn A (1999) Immunohistochemical detection of somatostatin sst2a receptors in the lymphatic, smooth muscular, and peripheral nervous systems of the human gastrointestinal tract: facts and artifacts. *J Clin Endocrinol Metab* 84:2942-2950
36. Reubi JC, Schar JC, Waser B, Wenger S, Heppeler A, Schmitt JS, Macke HR (2000) Affinity profiles for human somatostatin receptor subtypes SST1-SST5 of somatostatin radiotracers selected for scintigraphic and radiotherapeutic use. *Eur J Nucl Med* 27:273-282
37. Reubi JC, Schaer JC, Wenger S, Hoeger C, Erchegyi J, Waser B, Rivier J (2000) SST3-selective potent peptidic somatostatin receptor antagonists. *Proc Natl Acad Sci USA* 97:13973-13978
38. Reubi JC, Waser B, Liu Q, Laissue JA, Schonbrunn A (2000) Subcellular distribution of somatostatin sst2A receptors in human tumors of the nervous and neuroendocrine systems: membranous versus intracellular location. *J Clin Endocrinol Metab* 85:3882-3891
39. Schaer JC, Waser B, Mengod G, Reubi JC (1997) Somatostatin receptor subtypes sst1, sst2, sst3 and sst5 expression in human pituitary, gastroentero-pancreatic and mammary tumors: comparison of mRNA analysis with receptor autoradiography. *Proc Natl Acad Sci USA* 94:1155-1160
40. Schindler M, Sellers LA, Humphrey PP, Emson PC (1997) Immunohistochemical localization of the somatostatin SST2(A) receptor in the rat brain and spinal cord. *Neuroscience* 76:225-240
41. Schindler M, Holloway S, Humphrey PP, Waldvogel H, Faull RL, Berger W, Emson PC (1998) Localization of the somatostatin sst2(a)-receptor in human cerebral cortex, hippocampus and cerebellum. *Neuroreport* 9:521-525
42. Schonbrunn A (1999) Somatostatin receptors present knowledge and future directions. *Ann Oncol* 10:S17-S21
43. Schulz S, Schulz S, Schmitt J, Wiborny D, Schmidt H, Olbricht S, Weise W, Roessner A, Gramsch C, Holtt V (1998) Immunocytochemical detection of somatostatin receptors sst1, sst2A, sst2B, and sst3 in paraffin-embedded breast cancer tissue using subtype-specific antibodies. *Clin Cancer Res* 4:2047-2052
44. Sestini R, Orlando C, Peri A, Tricarico C, Pazzagli M, Serio M, Pagani A, Bussolati G, Granchi S, Maggi M (1996) Quantitation of somatostatin receptor type 2 gene expression in neuroblastoma cell lines and primary tumors using competitive reverse transcription-polymerase chain reaction. *Clin Cancer Res* 2:1757-1765
45. Slooter GD, Mearadji A, Breenman WA, Marquet RL, de Jong M, Krenning EP, van Eijck CH (2001) Somatostatin receptor imaging, therapy and new strategies in patients with neuroendocrine tumours. *Br J Surg* 88:31-40
46. Solcia E, Kloppel G, Sobin LH (2000) Histological typing of endocrine tumours. Springer, Berlin Heidelberg New York
47. Strowski MZ, Parmar RM, Blake AD, Schaeffer JM (2000) Somatostatin inhibits insulin and glucagon secretion via two receptors subtypes: an in vitro study of pancreatic islets from somatostatin receptor 2 knockout mice. *Endocrinology* 141:111-117
48. Volante M, Pecchioni C, Bussolati G (2000) Post-incubation heating significantly improves tyramide signal amplification. *J Histochem Cytochem* 48:1583-1585
49. Wulbrand U, Wied M, Zofel P, Goke B, Arnold R, Fehmann H (1998) Growth factor receptor expression in human gastroenteropancreatic neuroendocrine tumours. *Eur J Clin Invest* 28:1038-1049
50. Yamada Y, Post SR, Wang K, Tager HS, Bell GI, Seino S (1992) Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney. *Proc Natl Acad Sci USA* 89:251-255
51. Yamada Y, Reisine T, Law SF, Ihara Y, Kubota A, Kagimoto S, Seino M, Seino Y, Bell GI, Seino S (1992) Somatostatin receptors, an expanding gene family: cloning and functional characterization of human SSTR3, a protein coupled to

GLUT1 Messenger RNA and Protein Induction Relates to the Malignant Transformation of Cervical Cancer

Christian Rudlowski, MD,¹ Albert J. Becker, MD,² Willibald Schroder, MD,³ Werner Rath, MD,¹ Reinhard Büttner, MD,⁴ and Markus Moser, PhD⁵

Key Words: GLUT1; HPV; Human papillomavirus; Cervical neoplasm; Glucose metabolism

DOI: 10.13094/KYNOM5R62JW2GD7

Abstract

We studied whether induction of glucose transporters (GLUTs) 1 to 4 correlates with human papillomavirus (HPV)-dependent malignant transformation of cervical epithelium. Tissue samples of cervical intraepithelial neoplasia (CIN; grades 1 to 3), invasive carcinomas, and lymph node metastasis were examined. HPV typing was performed. Tissue sections were immunostained with GLUT1 to GLUT4 antibodies. Messenger RNA (mRNA) *in situ* hybridization confirmed GLUT1 protein expression.

Weak expression of GLUT1 was found in nondysplastic HPV-positive and HPV-negative epithelium; significant expression was observed in preneoplastic lesions, correlating with the degree of dysplasia. In CIN 3 high-risk HPV lesions, cervical cancer, and metastasis, GLUT1 was expressed at highest levels with a strong correlation of GLUT1 mRNA and protein expression. Immunostains for GLUT2 to GLUT4 were negative.

Cervical tumor cells respond to enhanced glucose utilization by up-regulation of GLUT1. The strong induction of GLUT1 mRNA and protein in HPV-positive CIN 3 lesions suggests GLUT1 overexpression as an early event in cervical neoplasia. GLUT1 is potentially relevant as a diagnostic tool and glucose metabolism as a therapeutic target in cervical cancer.

It has been known for several decades that tumor cells show enhanced glucose metabolism compared with benign tissues.¹ A continuous supply of glucose is the predominant source of adenosine triphosphate generation and substrate storage in mammalian cells. Molecular cloning of the glucose transporter protein 1 (GLUT1) that catalyzes the uptake of glucose in erythrocytes² and the subsequent identification of homologous genes (GLUT2, GLUT3, GLUT4, and GLUT5) revealed a family of genes to facilitate diffusion of hexoses into mammalian cells.³ On the basis of sequence similarities and functional characteristics, GLUTs can be divided into 3 subfamilies, referred to as class I (GLUTs 1-4), class II (GLUTs 5, 7, 9, and 11), and class III (GLUTs 6, 8, 10, and 12 and the myo-inositol transporter HMIT1).⁴

Altered expression of glucose transporter proteins has been described in different tissues under various conditions such as transformation by oncogenes, hypoxia, and exposure to insulin.⁵ *In vivo*, the storage of glycogen has been related to the differentiation status of squamous cell epithelia.^{6,7} In glucose-deprived 3T3-L1 adipocytes, GLUT1 activation by glucosylation relates to increased storage of glycogen.⁸ Cell culture experiments showed that the expression of GLUT1 transporter protein is induced by certain oncogenes such as *ras* and *src*,^{9,10} and regulated by the growth factors such as platelet-derived growth factor and epidermal growth factor.^{11,12}

Furthermore, members of the class I family of glucose transporter proteins were expressed strongly in various solid tumors such as breast cancer,^{13,14} renal cell carcinomas,^{15,16} brain tumors,¹⁷ and gastrointestinal malignomas.¹⁸ So far, class II and III GLUT members have not been observed to be relevant for tumorigenesis or tumor progression.¹⁹

Rudlowski et al / GLUT1 EXPRESSION IN CERVICAL NEOPLASMS

Little is known about the expression pattern of GLUTs in cervical neoplasms. Increased glucose uptake demonstrated by positron emission tomography (PET)²⁰ suggests a potential pathophysiologic role of glucose metabolism in cervical cancer. The malignant progression of cervical carcinomas from preinvasive lesions may serve as a paradigm for tumorigenesis along well-defined pathologic stages and pathogenesis. The development of cervical neoplasia is associated with the human papillomavirus (HPV). Viral integration guarantees perpetual expression of HPV-related viral oncogenes. These molecular changes are produced most consistently by the prototypic high-risk HPV types (eg, 16, 18, 31, and 33), but many others (low-risk types, eg, 6, 11, and 40) may be associated with invasive cancer.

We studied the relationship of HPV-related cervical tumor transformation, cellular glycogen storage and expression, and distribution of GLUT1, GLUT2, GLUT3, and GLUT4.

Materials and Methods

Tissue sections were prepared from formalin-fixed, paraffin-embedded archival specimens of cervical neoplasms obtained by surgery. Of 90 available tissue samples from cone biopsies performed for precancerous cervical lesions, 18 were mild cervical intraepithelial neoplasia (CIN 1), 24 were moderate (CIN 2), and 48 were severe dysplastic lesions (CIN 3). The cervical carcinoma specimens and the cervical lymph node metastasis specimens were obtained from 94 patients who underwent radical surgical resections at the Department of Gynecology and Obstetrics, University Hospital, Aachen, Germany, between August 1995 and September 1997. Tissue specimens from 22 pelvic lymph node metastases were studied. The histologic diagnosis and grade of dysplasia were confirmed by 2 pathologists (R.B. and M.M.) according to the criteria of the International Federation of Gynecology and Obstetrics (FIGO classification) and the World Health Organization.²¹ The tissue specimens were examined histologically and contained predominantly neoplastic cells, and the population of tumor cells was no less than 50% of all cells in the specimens. Paraffin-fixed tissue sections from 34 unselected benign cervical lesions were used as control samples and contained changes such as inflammatory or squamous atypia, immature metaplasia, or basal cell hyperplasia.

For immunohistochemical analysis, 3- μ m sections were deparaffinized and rehydrated. After rinsing with phosphate-buffered saline, endogenous peroxidase activity was inhibited by incubation in 3% hydrogen peroxide solution for 30 minutes, and the slides were incubated in a 1:10 solution of equine normal serum to block nonspecific staining.

After antigen retrieval by microwave treatment (600 W) for 5 minutes (3 times) in a 10-mmol/L concentration of sodium citrate, slides were incubated for 60 minutes with

polyclonal rabbit antihuman GLUT1 (DAKO, Hamburg, Germany) and polyclonal rabbit GLUT2, GLUT3, and GLUT4 antibodies (Biogenesis, New Field, England). The antibodies were diluted with bovine serum albumin according to the manufacturer's instructions. As a negative control, slides were reacted with equal amounts of nonspecific IgG instead of the specific primary antibody. After rinsing in phosphate-buffered saline, slides were incubated with a 1:50 dilution of biotinylated secondary antibody (DAKO) for 30 minutes and developed using the ABC technique (Vectastain ABC Systems, Vector Laboratories, Peterborough, England) according to the manufacturer's instructions. Immunocomplexes were visualized with the chromogen diaminobenzidine. Staining for glycogen was performed by the periodic acid-Schiff (PAS) reaction and counterstaining with toluidine blue.

The specificity of GLUT1 staining was confirmed routinely in all cases by staining of parallel tissue sections with GLUT1 antiserum that had been preincubated with the immunizing peptide (25 μ g/mL for 1 hour at 25°C). GLUT1 immunostaining was blocked by this peptide competition in all cases. The sections were evaluated semiquantitatively for immunoreactivity and PAS staining, based on a scoring system that was defined by the product of the intensity of the staining and the percentage of the stained cells. The specimens were subdivided into strong (50% or more marked cells), moderate (25%-49%), weak (10%-24%), and not marked for GLUT1 and glycogen (<10%). Grading was performed by consensus of 2 observers (R.B. and M.M.) in a blinded manner without previous knowledge of the histopathologic and clinical data for each case.

In situ hybridization was performed on paraffin-embedded tissue samples from different cervical biopsies as described previously²² using phosphorus 33 (³³P)-labeled sense and antisense complementary RNA riboprobes for rat GLUT1. Briefly, Proteinase K-pretreated slides (10 μ g/mL) were acetylated in acetic anhydride diluted 1:400 in a 0.1-mol/L concentration of triethanolamine (pH 8.0) and hybridized overnight in 50% formamide, 10% dextran sulfate, a 10-mmol/L concentration of tris(hydroxymethyl)aminomethane (pH 8.0), a 10-mmol/L concentration of sodium phosphate (pH 7.0), 2 \times standard saline citrate, a 5-mmol/L concentration of EDTA (pH 8.0), 150- μ g/mL concentration of transfer RNA, a 10-mmol/L concentration of dithiothreitol, and a 10-mmol/L concentration of β -mercaptoethanol supplemented with 5 \times 10⁴ cpm/ μ L of ³³P-labeled sense or antisense riboprobes at 50°C. Finally, slides were washed twice in 50% formamide-2 \times standard saline citrate-20-mmol/L concentration of β -mercaptoethanol, digested with RNase A (20 μ g/mL) for 30 minutes at 37°C, and washed again 3 times with the same washing buffer for 30 minutes each at 50°C. After dehydrating, slides were

coated with Kodak NTB2 emulsion (Kodak, Rochester, NY) and exposed for 8 to 10 days. The intensity of the silver staining was examined by darkfield microscopy, and digitalized images were examined using ImageJ 1.27z software (National Institutes of Health, Bethesda, MD). The Student *t* test was used for statistical analysis.

For detection and typing of HPV, tissue sections were digested by Proteinase K, and DNA was extracted using the DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. To detect HPV DNA, a 2-tiered polymerase chain reaction (PCR)-direct sequencing method was performed according to Feoli-Fonseca et al²¹ with slight modifications. We used the general consensus primers GP5+/GP6+ and the MY09/MY11 primers for amplification of HPV DNA. Forty cycles of amplification were run with an initial denaturation at 95°C for 5 minutes and for 30 seconds in each cycle. Temperatures of 37°C and 58°C were set for 30 seconds for annealing of the primer pairs GP5+/GP6+ and MY09/MY11, respectively. An extension step was done at 72°C for 5 minutes. The integrity of human genomic DNA was verified by PCR amplification of the β -globin gene. This reaction served as a positive control. The amplification products of the 2 consensus primer pairs and the β -globin PCR were run on a 2% agarose gel and stained with ethidium bromide.

PCR products were purified using the High Pure PCR product purification kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. The sequence of 1 strand of the purified PCR fragments was determined with the Big-Dye Terminator sequencing kit (Applied Biosystems, Foster City, CA) using 3 to 5 pmol of GP5+ or MY09 as the sequencing primers. The results of the sequencing reactions were analyzed on an ABI Prism 310 automated sequencer (Applied Biosystems). The obtained sequences were compared with documented virus sequences available in the GenBank databank using the BLAST program (Blast, Pittsboro, NC). The high-risk group comprised HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68, and the low-risk group included HPV types 6, 11, 40, 42, 43, and 44. The Student *t* test was used for statistical analysis.

Results

GLUT1-expressing epithelial cells were observed in squamous epithelial cells of normal and dysplastic cervical tissue probes. However, GLUT1 expression in normal epithelial cells was limited to the basal cell layers (Image 1A). Staining intensity was weak and varied in the basal compartment from a focal to homogeneous pattern distribution. All 34 specimens obtained from benign cervical tissue were classified as weak or not marked. Compared with GLUT1 expression in the proliferative cell compartment in the basal cell layer, changes such as inflammatory atypia,

metaplasia, and hyperplasia showed similar GLUT1 expression patterns and also were classified as weakly marked.

The intensity of GLUT1 expression was unaltered from the HPV status. Of 14 HPV-positive specimens, 6 showed high-risk positivity and 8 showed low-risk positivity (Table 1).

Erythrocytes and vascular endothelium served as internal positive controls and demonstrated strong GLUT1 positivity. Simultaneous PAS staining of the epithelium revealed the glycogen content to be correlated inversely to the expression of GLUT1 (Figure 1). The glycogen staining was limited predominantly to the upper cell compartments, which contain the functional epithelium. In good agreement with these immunostaining results, *in situ* hybridization showed GLUT1 messenger RNA (mRNA) expression to be present specifically in the basal epithelium (Image 1B).

In mildly dysplastic epithelium ($n = 18$), weak GLUT1 expression was limited to the basal cell layers (Image 1C) and (Image 1D), and in moderate dysplasia ($n = 24$), enhanced expression was extended to the basal and parabasal compartments. The GLUT1 expression pattern was independent of HPV status. In CIN 1 lesions, 4 cases were in the high-risk HPV group, 6 were in the low-risk HPV group, and 8 were HPV-negative cases. Compared with benign cervical changes, significantly increased GLUT1 expression was observed in CIN 1 lesions (Image 2).

Strong GLUT1 mRNA and protein expression was observed in all 48 CIN 3 lesions of the cervix uteri (Image 1E) and (Image 1F). GLUT1 expression was continuously strong throughout the entire dysplastic epithelium, including the basal, suprabasal, and apical parts of the cervical epithelium. Of 48 CIN 3 lesions, 46 were HPV positive; 34 cases belong to the high-risk group and 12 to the low-risk group. No differences between the HPV type or status and GLUT1 expression were observed. Only weak glycogen storage was observed in CIN 3 lesional tissue samples.

Tumors with moderate dysplasia showed a high percentage of HPV positivity (high-risk HPV, 16 cases; low-risk HPV, 3 cases; HPV type not assignable, 2 cases; HPV-negative, 3 cases). The intensity of GLUT1 protein expression depended on the extent of dysplasia and was accompanied by inversely correlated glycogen content in the nondysplastic areas of the epithelium (Figure 1).

GLUT1 overexpression was demonstrated in all 94 cases of invasive cervical carcinoma and all 22 cervical metastases (Image 1G) and (Image 1H). The intensity of GLUT1 protein expression did not vary significantly between different tumor stages, the degree of tumor differentiation, or HPV status. A high-risk HPV type was detectable in 84 cases and a low-risk type in 8 cases. In 2 specimens, no HPV was detectable.

GLUT1 expression was observed predominantly at the tumor periphery compared with the center of well-differentiated

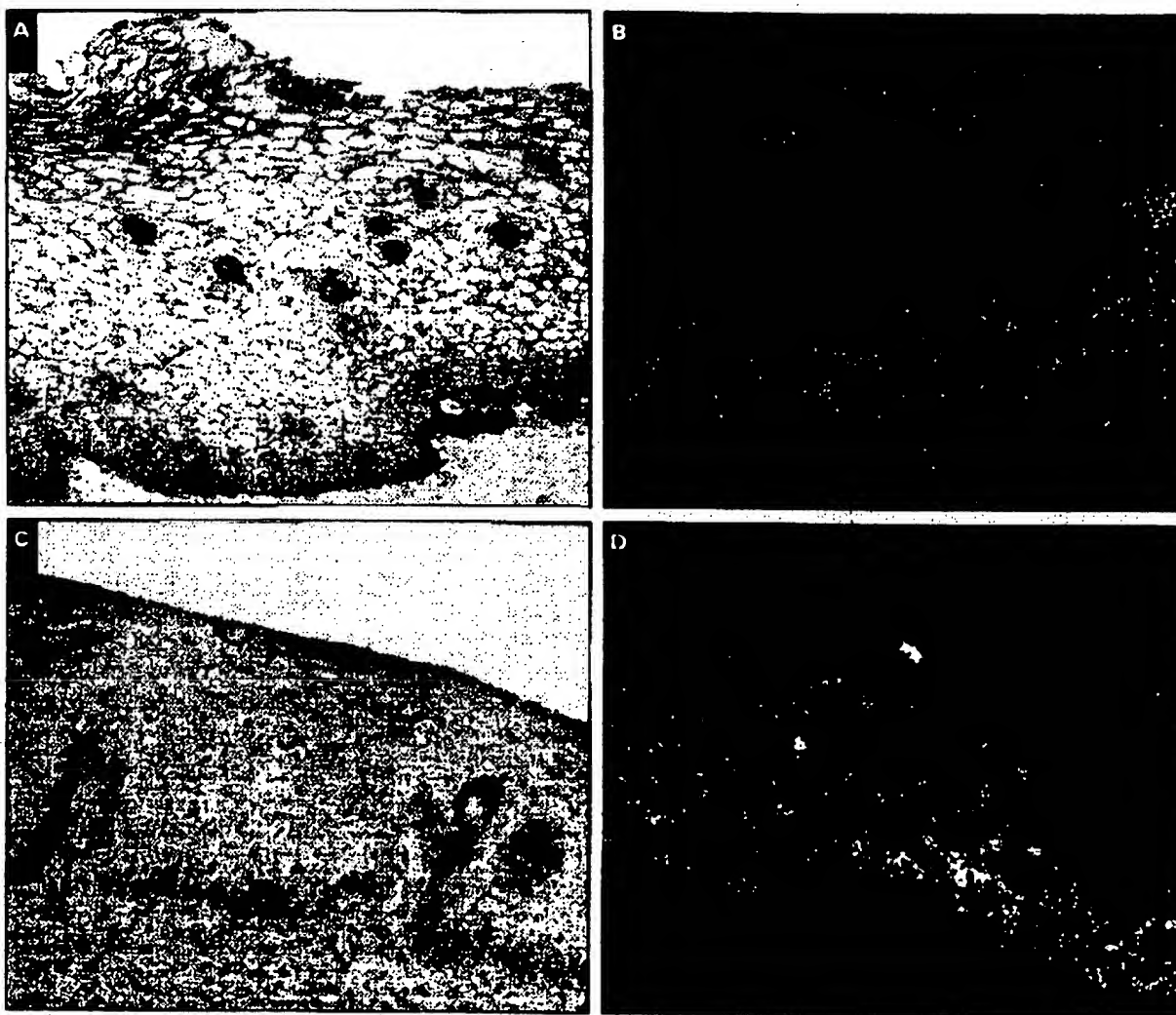


Figure 1 Glucose transporter (GLUT) 1 immunoreactivity with simultaneous periodic acid-Schiff (PAS) staining and GLUT1 messenger RNA (mRNA) silver staining in normal cervical epithelium (**A**, GLUT1 and PAS, $\times 200$; **B**, GLUT1 mRNA and silver staining, $\times 200$), cervical intraepithelial neoplasia (CIN) 1 lesions (**C**, GLUT1 and PAS, $\times 200$; **D**, GLUT1 mRNA and silver staining, $\times 200$).

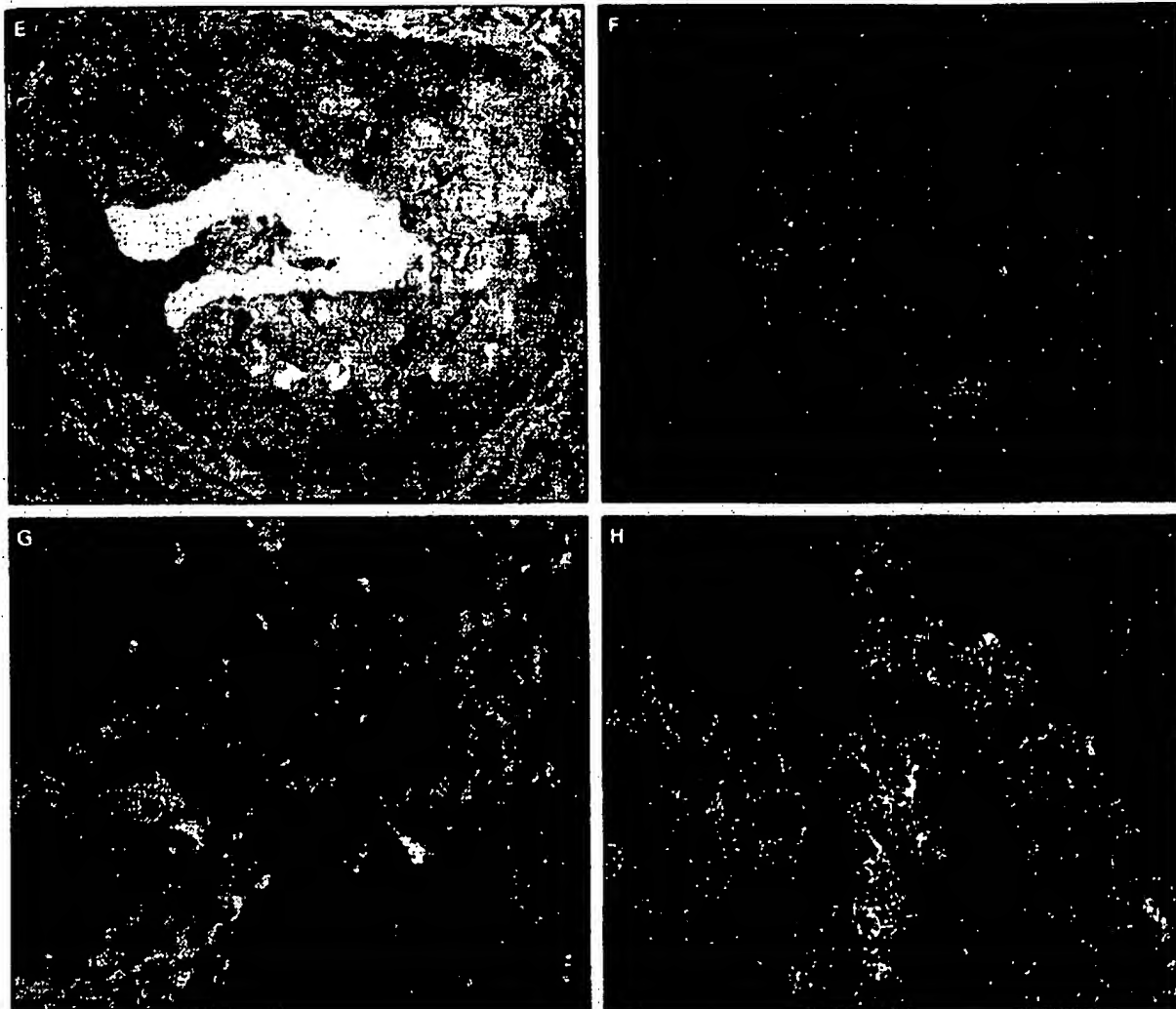
tumor areas. By PAS staining, we found an accumulation of glycogen in well-differentiated tumor areas, which was correlated inversely with GLUT1 expression. A similar distribution pattern was found in the lymph node metastases. **Figure 2** illustrates the results of GLUT1 mRNA in situ hybridization. The intensity of the specific silver staining correlated significantly with the extent of cervical dysplasia and showed similar levels in CIN 3 lesions and cervical carcinoma. Overexpression of mRNA was limited to dysplastic cells and matched the patterns of GLUT1 protein distribution revealed by immunohistochemical analysis.

Immunostains for GLUT2, GLUT3, and GLUT4 were negative in epithelial cells of both benign and dysplastic

lesions. Only inflammatory cells reacted with the antibodies and served as an internal positive control. In the advanced and poorly differentiated tumors, there was a pattern of expression similar to that in the dysplastic precursor lesions.

Discussion

Based on numerous molecular and epidemiologic studies, it has been demonstrated that specific types of HPV (high-risk HPVs, but also to a lower extent low-risk HPVs) can be detected in a transcriptionally active form in about 95% of cervical cancer biopsy specimens.²⁴ Although the high-risk HPV-encoded oncogenes E6 and E7 induce



x200), CIN 3 lesions (E, GLUT1 and PAS, x200; F, GLUT1 mRNA and silver staining, x200), and cervical cancer (G, GLUT1 and PAS, x200; H, GLUT1 mRNA and silver staining, x200). Expression of GLUT1 mRNA and protein in normal epithelial cells was limited to the basal proliferative cell layers, whereas PAS staining was restricted to the upper cell compartments. In the CIN 3 lesions and cervical carcinoma, strong expression of GLUT1 mRNA and protein was observed.

Table 1
HPV Status, GLUT1 Expression, and PAS Staining in the Study Group*

	Benign Cervical Tissue (n = 34)	CIN 1 (n = 18)	CIN 2 (n = 24)	CIN 3 (n = 48)	Cervical Cancer (n = 94)
HPV positive	14 (41)	10 (56)	21 (88) [†]	46 (96)	92 (98)
High risk	6 (18)	4 (22)	16 (67)	34 (71)	84 (89)
Low risk	8 (24)	6 (33)	3 (13)	12 (25)	8 (9)
GLUT1 expression	+	+	++	+++	+++
PAS staining	+++	+++	++	+	-

CIN, cervical intraepithelial neoplasia; GLUT, glucose transporter; HPV, human papillomavirus; PAS, periodic acid-Schiff; -, negative; +, weak; ++, moderate; +++, strong.
* Data for HPV are given as number (percentage). High-risk HPV infection was related closely to the degree of dysplasia and coincided with increasing GLUT1 expression.
† In 2 cases (8%), the HPV type was not assignable.

Radlowski et al / GLUT1 EXPRESSION IN CERVICAL NEOPLASMS

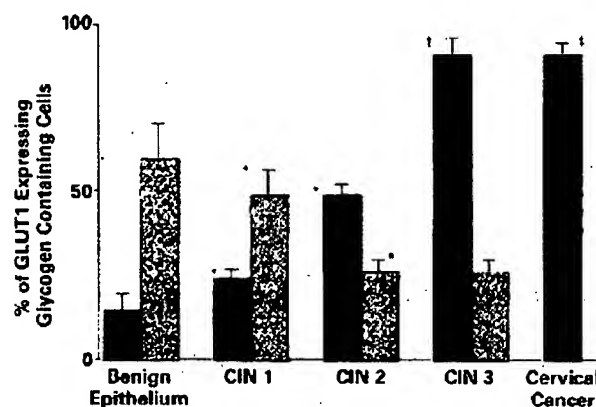


Figure 1 Glucose transporter (GLUT) 1 protein expression (black bars) and cellular glycogen content (gray bars) in benign and dysplastic cervical epithelium. The immunoreactivity of GLUT1 was associated strongly with the degree of dysplasia and correlated inversely with the cellular glycogen content. However, cervical intraepithelial neoplasia (CIN) 3 and cervical cancer showed equivalent levels of strong GLUT1 overexpression. * $P < .01$. † $P < .001$. ‡ In the center of well-differentiated cervical cancer, mild periodic acid-Schiff-positive staining was observed.

cellular immortalization,²⁵ their expression alone is not sufficient to cause malignant transformation.²⁶ Additional events must accumulate to convert a cell toward malignancy and to induce cervical cancer. It has been shown that the course of malignant disease not only is the result of failing intracellular

and immunologic surveillance mechanisms²⁷ but also is determined by efficient blood vessel formation to support the outgrowing tumor with nutrients, hormones, and oxygen.²⁸ Many tumorigenic cells are characterized not only by increased mitotic rates but also by high catabolic utilization of glucose and an increased number of specific glucose transporters.^{1,29} A sufficient intracellular glucose supply mediated by glucose transporters seems to be a prerequisite for enhanced HPV transcription and the expression of the E6 and E7 oncoproteins. In cell culture experiments using cervical cancer cells, specific suppression of the HPV 18 transcription was found by adding the glycolytic pathway inhibitor 2-deoxyglucose.³⁰

Our results demonstrate that human cervical carcinomas and their metastases selectively overexpress GLUT1. In benign cervical epithelium, moderate GLUT1 expression was restricted to the proliferating basal cell layers of the mucosa and changes such as inflammatory or squamous atypia, immature metaplasia, or basal cell hyperplasia. These findings are in line with the differentiation-related expression pattern of GLUT1 in normal human epidermis, in which GLUT1 is expressed in the basal layer and, to a lesser extent, in the immediate suprabasal layer of the epidermis.³¹ GLUT1 mRNA expression precisely coincided with protein overexpression in normal epithelium, cervical dysplasia, and invasive cancers. These findings strongly suggest that the increased GLUT1 expression in cervical neoplasia is mediated by transcriptional mechanisms.

In vitro studies on fibroblasts transfected with *ras* and *src* oncogenes revealed elevated levels of GLUT1 expression and an increase of transmembranous glucose transport.¹⁰ These

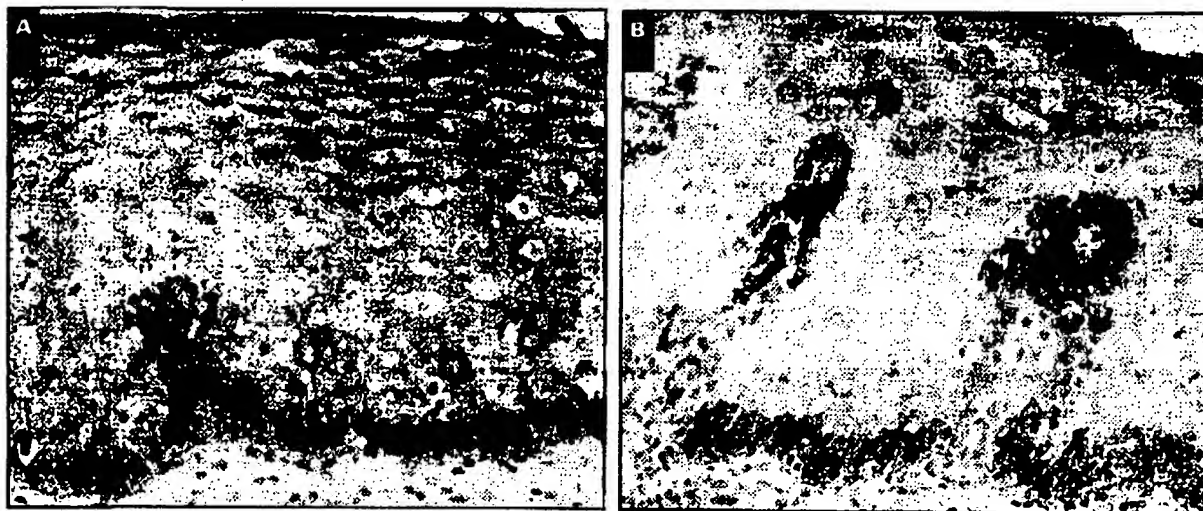


Image 2 Glucose transporter (GLUT) 1 protein immunoreactivity in cervical intraepithelial neoplasia 1 dysplasia (A, x200) vs reparative cervical changes (B, x200). The staining intensity of GLUT1 in low-grade dysplasia was increased markedly compared with the benign specimen. B, The arrow shows GLUT1 staining of erythrocytes, which serve as an internal control for GLUT1 positivity.

results support our present findings indicating that GLUT1 overexpression parallels malignant transformation of cervical epithelial cells. However, the regulatory mechanisms of GLUT1 induction are, so far, poorly understood. Tumor hypoxia might be a key step for up-regulation of GLUT1 transcripts.³²

However, we have observed overexpression in preneoplastic lesions. This might suggest direct oncogene-triggered mechanisms as potential factors for GLUT1 up-regulation. GLUT1 expression was not correlated with the HPV status or type in either benign or dysplastic cervical epithelium. High-risk HPV types were detectable predominantly in CIN 2 and 3 lesions and invasive cervical cancer and had good correlation with the dysplastic stage. All high-risk HPV-infected CIN 2 and 3 lesions and carcinoma showed enhanced GLUT1 expression. There is certain evidence that these cases are associated with virus transcription and the expression of E6 and E7 oncogenes.²⁴ An enhanced cellular glucose supply mediated by an increasing number of GLUT1 transporters could be a requirement for successful HPV-dependent malignant transformation.

Our results indicate that the clinical observation of an increased glucose uptake of cervical cancer cells was related to an exclusive transmembranous overexpression of GLUT1. These data strongly suggest that the observed relationship between increased GLUT1 expression and the intracellular accumulation of 2-fluoro-deoxy-D-glucose (FDG) in tumor cells represents the biologic basis for diagnosis using fluorine 18-labeled FDG PET.^{20,33} Enhanced expression of GLUT1 was detected already in moderate cervical dysplasia and much more in carcinoma in situ. In these preneoplastic lesions, the staining pattern of GLUT1 expression was stronger than in normal or mildly dysplastic mucosa and extended into the superficial cell layers depending on the degree of dysplasia. The observation that the preneoplastic lesions with increased GLUT1 expression were not detected by PET analysis may be explained by the small dimension of the epithelial dysplasia and the limited resolution of PET.

We also analyzed the cellular glycogen content to determine whether alterations in glucose uptake are linked to tumor-associated changes in the storage of this polysaccharide. The glycogen content was of special interest because the presence of glycogen is known to be related to cellular maturation of squamous epithelium and disappears with loss of differentiation during neoplastic transformation.⁶ In the present study, glycogen was found only in nondividing cells of the superficial layers of the benign cervical epithelium, whereas the basal layers did not contain glycogen (Image 1A). The glycogen content in normal and preneoplastic lesions generally correlated inversely with GLUT1 expression, exhibiting intense glycogen storage in normal and reduced or absent storage in dysplastic epithelium. Thus, reduction of glycogen content was related closely to the

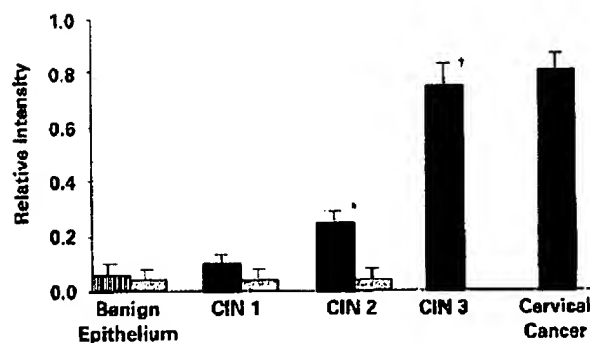


Figure 2 Glucose transporter (GLUT) 1 messenger RNA (mRNA) expression in the proliferative compartment of the benign cervical epithelium (hatched bar) and in dysplastic epithelium (black bars). The gray bars represent the relative intensity in the functional cell compartments. GLUT1 mRNA expression in cervical epithelium correlated significantly with the degree of dysplasia. * $P < .01$. † $P < .001$. CIN, cervical intraepithelial neoplasia.

degree of dysplasia and was associated strongly with increased GLUT1 expression.

The observed storage pattern of glycogen was correlated inversely with the extension of the proliferation compartment and is in line with previous observations showing an association between glycogen content and tumor cell development, as well as differentiation in squamous cell carcinoma tumorigenesis.³⁴ This cellular feature reflects the growing and metastasizing potency of cervical carcinomas, as increased utilization of glucose is a necessary condition of growing tumors and may be the pathophysiologic basis for induction of glucose transporter protein expression in the cell membrane. The lack of expression in benign cervical tissue and in the well-differentiated center of some cervical tumors gives rise to the assumption that GLUT1 expression might be regarded as a characteristic feature of tumor cell clones with an increased energy requirement.

In contrast with data from colon³⁵ and non-small cell pulmonary carcinoma,³⁶ no correlation between GLUT1 expression and prognostic clinicopathologic features was found in the present study. This difference can be explained easily, since significant overexpression of GLUT1 is observed in premalignant cervical lesions and well-differentiated carcinomas. We, therefore, believe that GLUT1 overexpression is an early event in the development of cervical carcinomas. Our findings also may have substantial implications for the potential therapeutic application of cytotoxic agents conjugated to glucose in the treatment of premalignant or malignant cervical lesions. Further studies are required to explore whether this strategy will lead to more specific drug delivery to the neoplastic cell compartment.

Rudlowski et al / GLUT1 EXPRESSION IN CERVICAL NEOPLASMS

From the ¹Department of Gynecology and Obstetrics, University Hospital Heidelberg, Heidelberg; ²Department of Neuropathology and ³Institute of Pathology, University of Bonn Medical Center, Bonn; ⁴Central Hospital, Bremen, University of Bonn Medical Center; and ⁵Department for Molecular Medicine, Max-Planck-Institut für Biochemie, Martinsried, Germany.

Address reprint requests to Dr Rudlowski: Dept of Gynecology and Obstetrics, University Hospital Heidelberg, Voßstr 7-9, D-69115 Heidelberg, Germany.

References

1. Warburg O. On the origin of cancer cells. *Science*. 1956;123:309-314.
2. Mueckler M, Canisio C, Baldwin SA, et al. Sequence and structure of a human glucose transporter. *Science*. 1985;229:941-945.
3. Gould GW, Bell GI. Facilitative glucose transporters: an expanding family. *Trends Biochem Sci*. 1990;15:18-23.
4. Joost HG, Bell GI, Best JD, et al. Nomenclature of the GLUT/SLC2A family of sugar/polyol transport facilitators. *Am J Physiol Endocrinol Metab*. 2002;282:E974-E976.
5. Olson AL, Pessin JE. Structure, function, and regulation of the mammalian facilitative glucose transporter gene family. *Annu Rev Nutr*. 1996;16:235-256.
6. Das DK, Chowdury JR. The use of glycogen studies in the evaluation of treatment for carcinoma of the cervix uteri. *Acta Cytol*. 1981;25:566-571.
7. Reisser C, Eichhorn K, Herndl-Mende C, et al. Expression of facilitative glucose transport proteins during development of squamous cell carcinomas of the head and neck. *Int J Cancer*. 1999;80:194-198.
8. McMahon RJ, Frost SC. Glycogen: a carbohydrate source for GLUT-1 glycosylation during glucose deprivation of 3T3-L1 adipocytes. *Am J Physiol*. 1996;270:E640-E645.
9. Birnbaum MJ, Haspel HC, Rosen OM. Transformation of rat fibroblasts by FSV rapidly increases glucose transporter gene transcription. *Science*. 1987;235:1495-1498.
10. Flier JS, Mueckler MM, Usher P, et al. Elevated levels of glucose transport and transporter messenger RNA are induced by *ras* or *src* oncogenes. *Science*. 1987;235:1492-1495.
11. Rollins BJ, Morrison ED, Usher P, et al. Platelet-derived growth factor regulates glucose transporter expression. *J Biol Chem*. 1988;263:16523-16526.
12. Hiraki Y, Rosen OM, Birnbaum MJ. Growth factors rapidly induce expression of the glucose transporter gene. *J Biol Chem*. 1988;263:13655-13662.
13. Alo PL, Visca P, Botti C, et al. Immunohistochemical expression of human erythrocyte glucose transporter and fatty acid synthase in infiltrating breast carcinomas and adjacent typical/atypical hyperplastic or normal breast tissue. *Am J Clin Pathol*. 2001;116:129-134.
14. Brown RS, Wahl RL. Overexpression of Glut-1 glucose transporter in human breast cancer: an immunohistochemical study. *Cancer*. 1993;72:2979-2985.
15. Ahn YS, Zerban H, Grobholz R, et al. Sequential changes in glycogen content, expression of glucose transporters and enzymic patterns during development of clear/acidophilic cell tumors in rat kidney. *Carcinogenesis*. 1992;13:2329-2334.
16. Nagase Y, Takata K, Moriyama N, et al. Immunohistochemical localization of glucose transporters in human renal cell carcinoma. *J Urol*. 1995;153:798-801.
17. Tsukamoto H, Hannada Y, Wu D, et al. GLUT1 glucose transporter: differential gene transcription and mRNA binding to cytosolic and polysome proteins in brain and peripheral tissues. *Brain Res Mol Brain Res*. 1998;58:170-177.
18. Niguchi Y, Marat D, Saito A, et al. Expression of facilitative glucose transporters in gastric tumors. *Hepato-gastroenterology*. 1999;46:2683-2689.
19. Joost HG, Thorens B. The extended GLUT-family of sugar/polyol transport facilitators: nomenclature, sequence characteristics, and potential function of its novel members (review). *Mol Membr Biol*. 2001;18:247-256.
20. Sugawara Y, Eishrich A, Kosuda S, et al. Evaluation of FDX PET in patients with cervical cancer. *J Nucl Med*. 1999;40:1125-1131.
21. Cervix uteri. In: *American Joint Committee on Cancer: AJCC Cancer Staging Manual*. 5th ed. Philadelphia, PA: Lippincott; 1997:189-194.
22. Moser M, Imhof A, Pscherer A, et al. Cloning and characterization of a second AP-2 transcription factor: AP-2 beta. *Development*. 1995;121:2779-2788.
23. Feoli-Fonseca JC, Oligny LL, Brochu P, et al. Human papillomavirus (HPV) study of 691 pathological specimens from Quebec by PCR-direct sequencing approach. *J Med Virol*. 2001;63:284-292.
24. zur Hausen H. Papillomavirus infections: a major cause of human cancers. *Biochim Biophys Acta*. 1996;1288:F55-F78.
25. Munger K, Phelps WC, Bubb V, et al. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol*. 1989;63:4417-4421.
26. Dürst M, Dzarlieva-Petrusevska RT, Boukamp P, et al. Molecular and cytogenetic analysis of immortalized human primary keratinocytes obtained after transfection with human papillomavirus type 16 DNA. *Oncogene*. 1987;1:251-256.
27. Altmann A, Jochmus I, Rosl F. Intra- and extracellular control mechanisms of human papillomavirus infection. *Intervirology*. 1994;37:180-188.
28. Risau W. Mechanisms of angiogenesis. *Nature*. 1997;386:671-674.
29. Younes M, Lechago LV, Somoano JR, et al. Wide expression of the human erythrocyte glucose transporter Glut1 in human cancers. *Cancer Res*. 1996;56:1164-1167.
30. Maehama T, Patzelt A, Lengert M, et al. Selective down-regulation of human papillomavirus transcription by 2-deoxyglucose. *Int J Cancer*. 1998;76:639-646.
31. Gherzi R, Melioli G, de Luca M, et al. "HepG2/erythroid/brain" type glucose transporter (GLUT1) is highly expressed in human epidermis: keratinocyte differentiation affects GLUT1 levels in reconstituted epidermis. *J Cell Physiol*. 1992;150:463-474.
32. Airley R, Lancaster J, Davidson S, et al. Glucose transporter glut-1 expression correlates with tumor hypoxia and predicts metastasis-free survival in advanced carcinoma of the cervix. *Clin Cancer Res*. 2001;7:928-934.
33. Umesaki N, Tanaka T, Miyama M, et al. The role of ¹⁸F-fluoro-2-deoxy-D-glucose positron emission tomography (¹⁸F-FDG-PET) in the diagnosis of recurrence and lymph node metastasis of cervical cancer. *Oncol Rep*. 2000;7:1261-1264.
34. Voldstedlund M, Dabelsteen E. Expression of GLUT1 in stratified squamous epithelia and oral carcinoma from humans and rats. *APMIS*. 1997;105:537-545.
35. Haber RS, Rathen A, Weiser KR, et al. GLUT1 glucose transporter expression in colorectal carcinoma: a marker for poor prognosis. *Cancer*. 1998;83:34-40.
36. Younes M, Brown RW, Stephenson M, et al. Overexpression of Glut1 and Glut3 in stage I nonsmall cell lung carcinoma is associated with poor survival. *Cancer*. 1997;80:1046-1051.



PERGAMON

European Journal of Cancer 39 (2003) 691-697

European
Journal of
Cancer

www.ejconline.com

Expression of deoxycytidine kinase in leukaemic cells compared with solid tumour cell lines, liver metastases and normal liver

C.L. van der Wilt^{a,1}, J.R. Kroep^a, W.J.P. Loves^a, M.G. Rots^{a,b,2},
C.J. Van Groeningen^a, G.J. Kaspers^b, G.J. Peters^{a,*}

^aDepartment of Medical Oncology, VU University Medical Center, Amsterdam, The Netherlands

^bDepartment of Pediatric Haematology/Oncology, VU University Medical Center, Amsterdam, The Netherlands

Received 20 May 2002; received in revised form 21 October 2002; accepted 3 December 2002

Abstract

Deoxycytidine kinase (dCK) is required for the phosphorylation of several deoxyribonucleoside analogues that are widely employed as chemotherapeutic agents. Examples include cytosine arabinoside (Ara-C) and 2-chlorodeoxyadenosine (CdA) in the treatment of acute myeloid leukaemia (AML) and gemcitabine to treat solid tumours. In this study, expression of dCK mRNA was measured by a competitive template reverse transcriptase polymerase chain reaction (CT RT-PCR) in seven cell lines of different histological origin, 16 childhood and adult AML samples, 10 human liver samples and 11 human liver metastases of colorectal cancer origin. The enzyme activity and protein expression levels of dCK in the cell lines were closely related to the mRNA expression levels ($r=0.75$, $P=0.026$ and $r=0.86$, $P=0.007$). In AML samples, dCK mRNA expression ranged from 1.16 to 35.25 ($\times 10^{-3} \times dCK/\beta\text{-actin}$). In the cell line panel, the range was 2.97–56.9 ($\times 10^{-3} \times dCK/\beta\text{-actin}$) of dCK mRNA expression. The enzyme activity in liver metastases was correlated to dCK mRNA expression ($r=0.497$, $P=0.05$). In the liver samples, these were not correlated. dCK mRNA expression showed only a 36-fold range in liver while a 150-fold range was observed in the liver metastases. In addition, dCK activity and mean mRNA levels were 2.5-fold higher in the metastases than in the liver samples. Since dCK is associated with the sensitivity to deoxynucleoside analogues and because of the good correlation between the different dCK measurements in malignant cells and tumours, the CT-RT PCR assay will be useful in the selection of patients that can be treated with deoxycytidine analogues.

© 2003 Published by Elsevier Science Ltd.

Keywords: Deoxycytidine kinase; Solid tumours; AML; Gemcitabine; Ara-C

1. Introduction

Deoxycytidine kinase (dCK) (EC 2.7.1.74) is a pyrimidine salvage enzyme that phosphorylates deoxycytidine, deoxyadenosine and deoxyguanosine. Furthermore, it is responsible for the phosphorylation of several deoxynucleoside analogues, which are widely used as anticancer and antiviral agents, such as 1- β -D-arabinosyl cytosine (cytarabine, Ara-C), 2'-2'-difluorodeoxycytidine (gemcitabine, dFdC) and

2-chlorodeoxyadenosine (cladribine, CdA). The activity of dCK shows a wide variation in normal and malignant cells and tissues [1,2].

The level of dCK activity has been reported to be closely related to sensitivity to deoxynucleoside analogues such as CdA [3], Ara-C and dFdC [4–9]. A deficiency of dCK leads to a resistance to deoxycytidine and deoxyadenosine analogues [10,11], therefore pretreatment measurements of dCK might be of predictive value [5]. However, relative large cell numbers or tumour biopsy specimens are required for measurements of activity. The use of the polymerase-chain reaction (PCR) enables us to use small biopsy specimens or a small number of cells. Therefore, we applied reverse transcriptase (RT)-PCR techniques to develop a sensitive assay for the expression of dCK mRNA. In order to quantify the expression, we used competitive templates

* Corresponding author. Tel.: +31-20-444-2633; fax: +31-20-444-3844.

E-mail address: gj.peters@vumc.nl (G.J. Peters).

¹ Present address: Comprehensive Cancer Centre (IKA), Amsterdam, The Netherlands.

² Present address: University Centre for Pharmacy, University of Groningen, The Netherlands.

(CTs) for *dCK* mRNA which were included in the same reaction mixture.

In this study, *dCK* mRNA expression was compared with *dCK* activity and protein expression in a panel of solid tumour and leukaemia cell lines, as well as in samples from adult and paediatric leukaemia patients, and from liver and liver metastases from patients with colorectal cancer.

2. Materials and methods

2.1. Materials

[8-³H]-2-Chloro-2'-deoxyadenosine (24.2 Ci/mmol) was obtained from Moravek, Brea, CA, USA. RNazol was purchased from Campro Scientific (Veenendaal, The Netherlands); Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV-RT) and RNase inhibitor (25 IU/μl) were obtained from Promega (Madison, WI, USA). Deoxyribonucleotides (dNTPs), random hexamers, and Taq polymerase (5 IU/μl) were purchased from Amersham/Pharmacia Biotech (Roosendaal, The Netherlands). Primers were from Isogen (Maarssen, The Netherlands). RESponse Research agarose was obtained from Biozym (Landgraaf, The Netherlands). The rabbit anti-human *dCK* polyclonal antibody was produced and tested at the Institute of Molecular Biology and Biotechnology at Crete by the group of Dr I. Talianidis [12].

2.2. Cells

The cancer cell lines BxPC3 (pancreas), UMSCC14C (head and neck) A2780 (ovarian), H322 (non-small cell lung), CCRF CEM (T-cell acute lymphoid leukaemia), U937 (histiocytic lymphoma) and HL60 (acute myeloid leukaemia (AML)) were used to measure *dCK* expression and activity. The variants AG6000 (A2780 resistant to gemcitabine), CCRF CEM-*dCK*⁻ (CCRF CEM resistant to AraC) were included to set the lowest level of *dCK* activity. The solid tumour cell lines were cultured in Dulbecco's modification of Eagles medium (Bio Whittaker Europe, Verviers, Belgium) supplemented with 5% fetal calf serum (FCS) (Gibco, Paisly, UK). The others were cultured in Roswell Park Memorial Institute (RPMI) medium (Bio Whittaker) with 10% FCS. Subconfluent flasks with cells were harvested and counted. The cell pellets were frozen at -70 °C until further analysis.

2.3. Acute myeloid leukaemia samples

Mononuclear blast cells were derived after informed consent from 4 adult and 12 paediatric patients with acute myeloid leukaemia as previously described in Ref.

[13] and isolated by density gradient methods. Contaminating non-leukaemic cells were removed by immunomagnetic beads resulting in a percentage of leukaemic cells above 80%, as morphologically determined by May-Grünwald-Giemsa staining of the cytopins.

2.4. Liver and liver metastases

Biopsies of liver metastasis and normal liver of patients with histologically-proven colorectal cancer were used to study *dCK* in solid tumours and normal tissue. Samples were immediately frozen in liquid nitrogen and subsequently stored at -80 °C. Frozen tissues were pulverised using a microdismembrator [1], powdered tissue could be stored for the different *dCK* assays.

2.5. Deoxycytidine kinase activity

dCK activity in each of the cell lines was determined using ³H-CdA as a substrate, since CdA is a better and more specific substrate than deoxycytidine, which can also be phosphorylated by thymidine kinase 2. The assay was based on a method described by Arnér and colleagues in Ref. [14]. Cell pellets (at least 1.5×10⁷) were resuspended in 50 mM Tris(hydroxymethyl)-aminomethane (TRIS)/HCl buffer (pH 7.4), containing 4 mM dithiothreitol (DTT), at a concentration of 30×10⁶ cells/ml and kept on ice. Frozen tissue was suspended in 50 mM TRIS/HCl buffer (0.33 g/ml). Cells were disrupted by sonification and the suspension was centrifuged for 15 min at 7000g. The supernatant was used for the enzyme assay and protein determination. We used 25 μl of supernatant (6-126 μg protein) and added 10 μl 50 mM MgATP/100 mM NaF, 10 μl 250 μM [8-³H]-CdA (specific activity 0.19 μCi/nmol) and 5 μl TRIS/HCl buffer. This mixture was incubated at 37 °C. The assay was stopped by heating the samples to 90 °C, then 10 μl of 10 mM CdA was added to enable its localisation by ultraviolet (UV) absorbance on the PEI thin layer sheets. The samples were centrifuged in order to spin down the protein and subsequently 5 μl was spotted onto polyethyleneimine (PEI) cellulose thin-layer chromatography sheets. The sheets were developed with water to separate substrate and product. The different spots were cut and put into a vial with 750 μl of 0.1 M HCl/0.2 M KCl to enable elution of radioactivity from the PEI cellulose, during 2 h of shaking. After addition of liquid scintillation fluid, radioactivity was counted. Enzyme activity was expressed as pmol CdA-MP formed per h per mg protein.

2.6. Competitive template RT-PCR

The isolation of RNA was performed with RNazol as previously described in Ref. [13]. Cell pellets were

suspended in an aliquot of 200 μ l of RNazol per 10⁶ cells. Twenty micrograms of the isolated RNA was used for reverse transcription into cDNA. Random hexamers were used as primers for M-MLV-RT at a concentration of 0.045 μ g/ μ l. After a brief incubation at 56 °C to remove secondary structures, samples were quickly cooled down on ice and annealing of the hexamers also took place on ice. The extension of the primers was at 42 °C using M-MLV-RT. The reaction was terminated by heating at 95 °C for 5 min. cDNA samples were stored at -20 °C until further use.

The design of the primers and the CT was based on the method described by Willey and colleagues in Refs. [13,15,16]. One forward (A) and two reverse primers (BC and C) were selected by Oligo software requiring an optimal annealing temperature of 58 °C, absence of hairpins and no predictable stable primer-dimer formations.

A: 5'-GAAGGGAACATCGCTGCAGG
C: 5'-TGGCCAAATTGGTTATTTCATCC
BC: 5'-TGGCCAAATTGGTTATTCATCCTTGAG-
CTTGCCATTGAGAGAGG

The primers A and C were used to amplify a native template of 425 bp covering exons 2, 3 and part of 4 of *dCK*. The primers A and BC were selected for the construction of the CT. The B sequence is upstream of C and amplification with A and BC resulted in a shorter product of 294 bp. This product was the CT linked by the A and C sequence, which was further purified and quantified as described in Refs. [13,15]. The β -actin gene was used as a reference gene and the previously published construction of the β -actin CT was used [16].

CTs of β -actin and *dCK* were mixed in a ratio of 10, 100 and 1000 (10^{-12} M/ 10^{-13} M; 10^{-12} M/ 10^{-14} M; 10^{-12} M/ 10^{-15} M) based on preliminary experiments. One single master mix was prepared for every cDNA sample containing PCR buffer (1 \times), MgCl₂ (1.5 mM), dNTPs (200 μ M), Taq polymerase (5 μ l, final concentration 0.02 U/ μ l), sample cDNA and CT mix in a total volume of 46 μ l. *dCK* and β -actin primers (4 μ l, final concentration 4 ng/ μ l of each primer) were added to different tubes, which already contained certain aliquots of the master mix. Reaction mixtures were overlaid with mineral oil and cycled in a MJ Research PTC-200 (Biozym, Landgraaf, The Netherlands). A semi-hot start from ice to 94 °C was performed and with 1-min steps of denaturation at 94 °C, primer annealing at 58 °C and elongation at 72 °C the PCR reaction continued for 35 cycles.

PCR products were separated by electrophoresis on 2% agarose gels containing 0.1 μ g/ml ethidium bromide. Quantification was by digital image analysis using Scion Software. The concentration of native template molecules in the cDNA samples was calculated by the ratio of the intensity of native template/CT and the

molarity of the CT mixture as previously described in Refs. [13,15,16]. Normalisation of the *dCK* expression to the expression of β -actin was based on data obtained from assays with the same master mixture.

The *dCK* expression in the cell lines was determined at least three times on different cDNA samples of one cell line. The expression in the AML samples was measured once, because these samples were also used for several other assays. Levels of expression were reported as units *dCK* mRNA β -actin mRNA $\times 10^{-3}$ molecules.

2.7. Western blotting

Western blots were produced according to a published method in Ref. [12]. Protein extracts (30 μ g) were electrophoresed on 12.5% sodium dodecyl sulphate (SDS)-polyacrylamide gels and transferred by electroblotting to nitrocellulose membranes (Hybond ECL membranes, Amersham, Buckinghamshire, UK). After overnight incubation in blocking buffer (5% bovine serum albumin in TRIS-buffered saline with 0.1% Tween 20), the blot was probed with rabbit anti-human *dCK* polyclonal antibody (1:5000, 1 h at room temperature) and goat anti-rabbit secondary antibody conjugated to horseradish peroxidase (1:40,000, 1 h at room temperature). Purified recombinant *dCK* (10 ng), extracts of AG6000 cells lacking *dCK* and parental A2780 ovarian cancer cells were tested on every blot to identify the *dCK* band. Levels of relative expression were determined by densitometric scanning (Imaging Densitometer, model GS-690, Biorad of the X-ray films (Hyperfilm ECL). *dCK* protein levels were expressed as ng *dCK* per mg total protein.

2.8. Statistics

The Spearman correlation test was used to quantify the relationship between *dCK* activity and *dCK* expression at the mRNA and protein level. Student's *t*-test was used for comparison of paired and unpaired samples.

3. Results

The panel of cell lines had a 10-fold variation in *dCK* activity. The lower level was set by CCRF-CEM/*dCK*⁻ and AG6000 (1.2 and 1.4 pmol/h per mg protein, respectively). The mRNA expression showed a 19-fold variation for the cell lines (Table 1). No *dCK* mRNA could be detected in CCRF-CEM/*dCK*⁻. AG6000 showed a normal and a truncated mRNA product as has been previously published [6], this complicated accurate quantification and this cell line was excluded from the mRNA expression evaluation (Fig. 1). The reproducibility of the CT-RT-PCR assay in the cell lines is given in Table 1 as the relative standard deviation.

Table 1
dCK mRNA expression in cell lines and AML samples^a

Sample	dCK mRNA expression	Relative standard deviation
Cell line	Ratio $10^{-3} \times dCK/\beta\text{-actin}$	(%)
CCRF CEM	38.0 ± 17.3	33
HL60	56.3 ± 19.9	26
U937	56.9 ± 6.60	7
A2780	6.06 ± 6.76	15
UMSCC14C	4.73 ± 0.38	10
BxPC3	2.97 ± 0.16	9
H322	5.97 ± 1.54	6
AML		
0438*	4.24	
4695*	1.87	
4779	3.12	
6250*	10.79	
7470	26.30	
7578	35.23	
7621	19.40	
7747	1.16	
7985	2.77	
8019	12.27	
8051	9.88	
8069	3.52	
8196	26.17	
8400*	11.39	
8773	3.00	
8774	4.40	

AML, acute myeloid leukaemia.

^a mRNA expression is the mean ± standard error of the mean (S.E.M.) of at least three measurements on different samples for the cell lines. The relative standard deviation (S.D.) is based on 2-3 measurements on the same sample. (*) Indicates adult AML samples, the others are from childhood AML; not enough clinical material was available to measure enzyme activity. No expression was measured in the Ara-C- and gemcitabine-resistant cell lines.

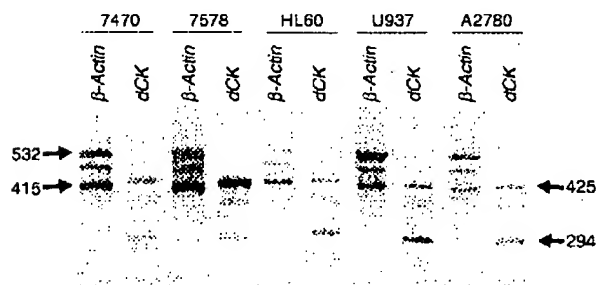


Fig. 1. Representative X-ray photos of β -actin and dCK gene and competitive template (CT) products of the cell lines and 2 AML samples (7470 and 7578). The blots show three bands for both dCK and β -actin: 532 and 425 are the number of the basepairs encoded by the forward and reverse primers for the native cDNA of β -actin and dCK, respectively; 415 and 294 are the number of basepairs encoded by the competitive templates for β -actin and dCK, respectively. The intermediate band is a heterodimer (HD) which can be formed between the native cDNA and the competitive template. The contribution of the HD was calculated as previously described in Refs. [13,16]. The bands were scanned and the optical density (OD) was used to calculate a ratio between the native cDNA and CT of each target, thereafter these ratios were used to calculate the ratio for dCK and β -actin.

The AML samples had a 30-fold range of dCK mRNA expression that overlapped with that of the cell lines, although the high level of expression in the cell lines HL60 and U937 was not seen in the patient samples.

Liver metastases contained a mean 2.5-fold higher dCK enzyme activity than the surrounding normal liver samples (10.1 ± 1.9 and 4.0 ± 0.4 nmol/h/mg protein, respectively, $P=0.004$, Student's *t*-test for paired data) (Fig. 2). The range of dCK activities in the liver metastases was about 17-fold (1.4–23.8 nmol/h per protein), while that in the liver ranged only 3-fold (2.0–6.7). The dCK mRNA expression showed a wide variation for these samples: a 150-fold difference (0.53–80.4). The subgroup of liver samples had a 36-fold range (0.63–22.7) in dCK mRNA expression levels.

The correlation between dCK activity and dCK mRNA levels in the cell lines is depicted in Fig. 3a. The correlation coefficient determined by Spearman rank correlation was $r=0.75$ ($P=0.026$). The correlation between

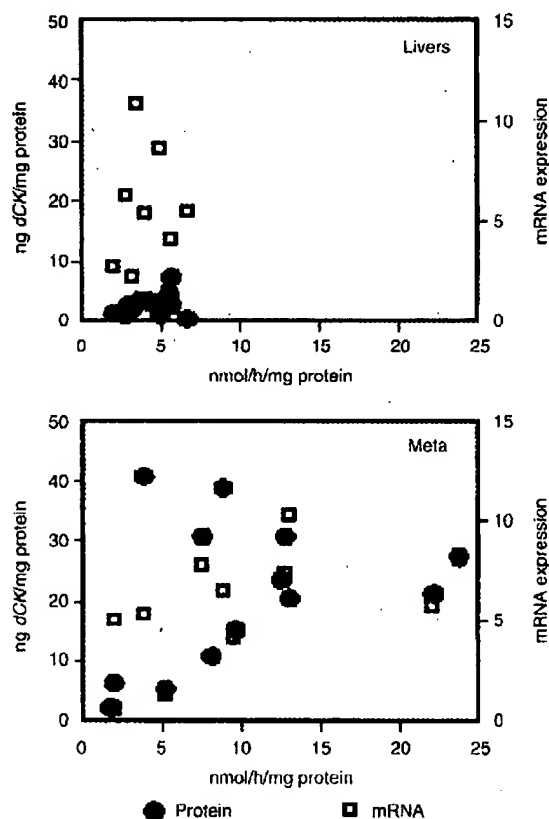


Fig. 2. Activity of dCK, amount of dCK protein and dCK mRNA expression in normal livers and liver metastases (meta) from colorectal cancer. The dCK activity in liver metastases was significantly higher than in livers ($P=0.004$). In livers, no correlation was found between dCK activity and dCK mRNA, while dCK activity was related to dCK protein ($r=0.59$; $P=0.013$). However, in liver metastases, dCK activity was correlated with dCK mRNA ($r=0.497$; $P=0.05$), but not to dCK protein. One liver metastasis sample is not shown having 72.9 μ g dCK/mg protein and 80.4 as dCK mRNA expression. One liver sample had a dCK mRNA of 22.7 and a dCK activity of 5 nmol/h/mg protein.

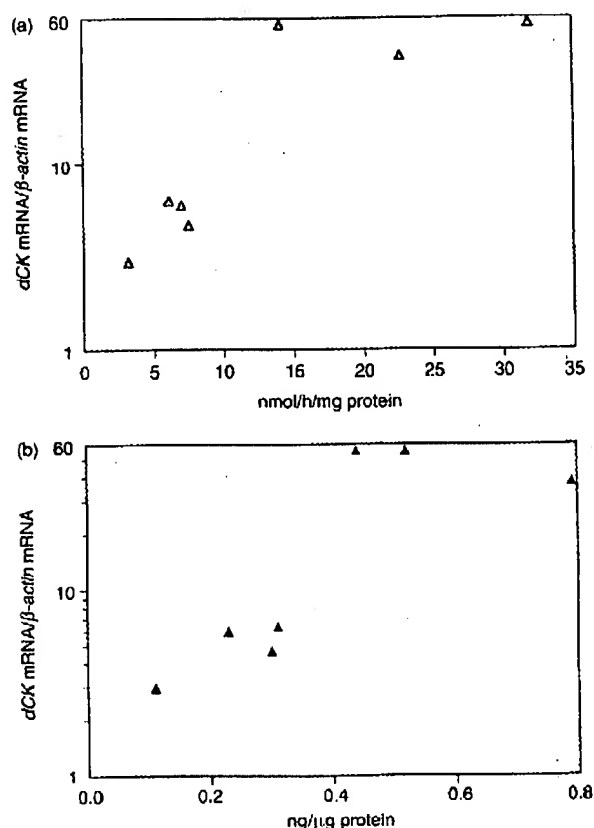


Fig. 3. (a) Correlation between mRNA expression and dCK activity in a panel of seven cell lines of different origin. Correlation coefficient $r = 0.75$ ($P = 0.026$). (b) Correlation between mRNA expression and dCK protein expression in a panel of seven cell lines of different origin. Correlation coefficient $r = 0.79$ ($P = 0.018$).

these dCK measurements in pooled liver and liver metastases is $r = 0.416$ ($P = 0.024$). The correlation was also observed for liver metastases only (Fig. 2) ($r = 0.497$, $P = 0.050$), but not for the subgroup of livers.

The dCK protein expression in the cell line panel (Figs. 3b and 4) varied about 7.2-fold. The correlation with dCK mRNA (Fig. 3b) was $r = 0.85$ ($P = 0.007$), as determined by Spearman rank correlation. In addition, the correlation coefficient of protein expression and activity was significant ($r = 0.79$, $P = 0.018$, Spearman).

In the liver metastases, dCK protein levels were clearly higher and had a broader range compared with the normal liver samples (5.3–72.9 and 0.5–7.4 ng dCK/mg protein, respectively) (Fig. 2). A correlation between

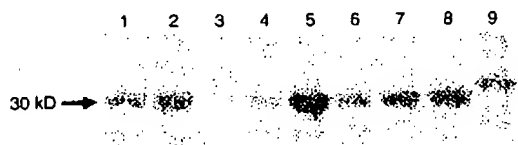


Fig. 4. Western blot of dCK protein in the different cell lines (1, 14C; 2, A2780; 3, AG6000; 4, BxPC3; 5, CEM; 6, H322; 7, HL60; 8, U937; 9, dCK-HIS).

dCK activity and protein was found in the liver samples ($r = 0.59$, $P = 0.013$), but not in the metastases. There was no correlation between dCK mRNA and protein expression in these samples.

4. Discussion

The expression of dCK mRNA in a panel of solid tumour cell lines and leukaemia and lymphoma cell lines was closely correlated to the activity of this enzyme. This relationship had already been established for a panel of six leukaemia cell lines by Kawasaki and colleagues in Ref. [17]. For solid tumour cells from different histological origin such a relationship had not yet been established. Any such correlation is of major interest because of the important role of some novel deoxynucleoside analogues, such as gemcitabine, in the treatment of solid tumours. Our panel is a mixture of both leukaemic and solid tumour cells, and if we considered only the four solid tumour cell lines, the relationship was weaker. The correlation between dCK activity and dCK mRNA expression in liver metastases was also weaker and its statistical significance was borderline ($P = 0.05$).

The substrate CdA, which we used, is also a substrate for deoxyguanosine kinase (dGK), albeit with a much higher K_m (85 μM) and lower efficiency than for dCK (5 μM) [18,19]. The cell lines CCRF-CEM-dCK⁻ and AG6000 have some CdA phosphorylating capacity, but they do not show a dCK protein band in western blotting. The activity measured here is most likely due to dGK activity, which is also present in the cell extract. The positive relationship between dCK enzyme activity and protein expression has also been observed in studies of Kawasaki [20] and Spasokoukotskaja [21]. The variation in the dCK mRNA expression in the AML samples (30-fold) is somewhat lower than that found in a study with 13 childhood AML samples (50-fold) [22] and 35 childhood ALL and AML samples [23] in which samples from relapsed leukaemic cells had a lower dCK expression. In addition, for the enzyme activities, a comparable range in dCK levels was found [24].

The relationship between dCK activity and response to treatment with deoxynucleoside analogues has been investigated in several malignancies. Hairy cell leukaemia, which is very sensitive to CdA showed both high dCK and dGK activity. The degree of total CdA phosphorylation correlated with the response to CdA treatment of both hairy cell leukaemias and chronic lymphocytic B-cell leukaemias, although other factors also seemed to be involved [3,25]. In AML, dCK activity correlated with the response to Ara-C treatment in 21 patients [26]. ALL patients more often relapsed when dCK expression was low or absent [27]. Transfection experiments with dCK and other deoxynucleoside kinase [28] genes has underlined the important role of

dCK in the antitumour activity of deoxynucleoside analogues. Increased expression of *dCK* gene by transfection restored and enhanced the sensitivity to deoxynucleoside analogues that are activated by dCK [8,9,29], both *in vitro* and *in vivo*.

Most of the AML samples were derived from childhood AMLs [13], which are always treated with an Ara-C-containing protocol. The large range of *dCK* mRNA expression, which overlapped with the expression of the cell lines, indicates that this disease in general has an intrinsic sensitivity to Ara-C, but that individual differences in sensitivity to Ara-C exist. Moreover, this has been reported and has prognostic significance [30,31]. Recently, a clinical protocol including Ara-C was started in which dCK expression will be evaluated prospectively.

Gemcitabine is currently being used for the treatment of many solid malignancies. Single agent gemcitabine shows little activity in colorectal cancer, but it is not clear whether this is due to a low dCK expression. Therefore, we designed this study to determine dCK expression in liver metastases. No information is currently available on the relationship between response and dCK activity in patients. In cell lines, we previously reported a relationship between gemcitabine sensitivity and dCK activity [6,7], while in xenografts [5], we also observed a relationship between gemcitabine sensitivity and dCK activity. *dCK* gene transfer enhanced the antitumour effect of gemcitabine on tumour xenografts [9], confirming the relationship between dCK and gemcitabine activity. It is remarkable that the dCK activity in the liver samples in our study showed relatively small variations and did not exceed a level of 5 nmol/h/mg protein. In contrast, protein and mRNA expression levels showed more variation. In liver metastases, both activity and mRNA varied considerably and correlated with one another. This would give an advantage for liver metastases compared with liver in selectively activating deoxynucleoside analogues in liver metastases. A high dose of gemcitabine is expected to lead to selective activation in metastases, and when given via a hepatic artery infusion would not leak to the systemic circulation as livers have a high deaminase activity. The much smaller range of dCK activity in the liver samples compared with the liver metastases also indicated that in the liver a posttranslational regulation mechanism exists, which seems to be deregulated in the liver metastases. It has been reported that dCK activity can be regulated by its phosphorylation [32,33]. dCK purified from leukaemic blasts can be phosphorylated by Protein Kinase C (PKC), resulting in a 100% increase in enzyme activity [32]. However, recombinant dCK is a poor substrate for PKC, but is effectively phosphorylated by protein kinase A, and this is not accompanied by an increase in enzyme activity [33]. An enhanced dCK activity could be eliminated by protein phosphatase treatment of lymphoid cells, supporting a secondary modification of the

dCK protein [34]. In addition, the promoter for *dCK* contains several potential regulatory transcription factor sites [35,36], although it is not yet clear whether this regulation is tissue-specific or deregulated in tumours.

For the enzyme activity assays and western blotting, relatively large pieces of tissue or many cells are required, and these are not always available. The current PCR assay detects a wide range of *dCK* mRNA levels. CT-RT-PCR can be performed with any PCR equipment, and does not need the requirement of an expensive Taq man or light cycler apparatus. In addition, results from this type of assay shows a high inter- and intralaboratory agreement [37]. Measurement of *dCK* mRNA expression offers the possibility for larger-scale studies, in patients with solid tumours and leukaemia, to determine a relationship between dCK expression and response to either gemcitabine or Ara-C. This data should be followed by a prospective study to select patients for deoxynucleoside analogue treatment based on their dCK activity.

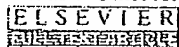
Acknowledgements

This work was supported by European Union Grant BIOMED BMH4-CT96-0479.

References

1. Ruiz van Haperen VWT, Veerman G, Braakhuis BJM, et al. Deoxycytidine kinase and deoxycytidine deaminase activities in human tumour xenografts. *Eur J Cancer* 1993; **29A**, 2132-2137.
2. Eriksson S, Arnér E, Spasokoukotskaja T, et al. Properties and levels of deoxynucleoside kinases in normal and tumor cells: implications for chemotherapy. *Adv Enzyme Regul* 1994; **34**, 13-25.
3. Kawasaki H, Carrera CJ, Piro LD, Saven A, Kipps TJ, Carson DA. Relationship of deoxycytidine kinase and cytoplasmic 5'-nucleotidase to the chemotherapeutic efficacy of 2-chlorodeoxyadenosine. *Blood* 1993; **81**, 597-601.
4. Kroep JR, Van Moersel CJ, Veerman G, et al. Role of deoxycytidine kinase, thymidine kinase 2, and deoxycytidine deaminase in the antitumor activity of gemcitabine. *Adv Exp Med Biol* 1998; **431**, 657-660.
5. Kroep JR, Loves WJP, Van der Wilt CL, et al. Pretreatment deoxycytidine kinase levels predict *in vivo* gemcitabine sensitivity. *Mol Cancer Ther* 2002; **1**, 371-376.
6. Ruiz van Haperen VWT, Veerman G, Eriksson S, et al. Development and molecular characterization of a 2',2'-difluorodeoxycytidine-resistant variant of the human ovarian carcinoma cell line A2780. *Cancer Res* 1994; **54**, 4138-4143.
7. Ruiz van Haperen VWT, Veerman G, Vermorken JB, Pinedo HM, Peters G. Regulation of phosphorylation of deoxycytidine and 2',2'-difluorodeoxycytidine (gemcitabine); effects of cytidine 5'-triphosphate and uridine 5'-triphosphate in relation to chemosensitivity for 2',2'-difluorodeoxycytidine. *Biochem Pharmacol* 1996; **51**, 911-918.
8. Hupke DM, Stegmann AP, Mitchell BS. Retroviral transfer of deoxycytidine kinase into tumor cell lines enhances nucleoside toxicity. *Cancer Res* 1996; **56**, 2343-2347.

9. Blackstock AW, Lightfoot H, Case LD, et al. Tumor uptake and elimination of 2', 2'-difluoro 2' deoxycytidine (gemcitabine) after deoxycytidine kinase gene transfer: correlation with in vivo tumor response. *Clin Cancer Res* 2001, 7, 3263-3268.
10. Peters GJ, Jansen G. Resistance to antimetabolites. In Schilsky RL, Milano GA, Ratain MJ, eds. *Principles of Antineoplastic Drug Development and Pharmacology*. New York, Marcel Dekker, 1996.
11. Bergman AM, Pinedo HM, Peters GJ. Determinants of resistance to 2', 2'-difluorodeoxycytidine (gemcitabine). *Drug Res Update* 2002, 5, 19-33.
12. Hatzis P, Said Al-Madhoon A, Jullig M, Petrakis TG, Eriksson S, Talianidis I. The intracellular localization of deoxycytidine kinase. *J Biol Chem* 1998, 273, 30239-30243.
13. Rots MG, Willey JC, Jansen G, et al. mRNA expression levels of methotrexate resistance related proteins in childhood leukemia as determined by a competitive template based RT-PCR method. *Leukemia* 2000, 14, 2166-2175.
14. Arnér E, Spasokoukotskaja T, Eriksson S. Selective assays for thymidine kinase 1 and 2 and deoxycytidine kinase and their activities in extracts from human cells and tissues. *Biochem Biophys Res Commun* 1992, 188, 712-718.
15. Willey JC, Coy EL, Frampton MS, et al. Quantitative RT-PCR measurement of cytochromes p450 1A1, 1B1, and 2b7, microsomal epoxide hydrolase and NADPH oxidoreductase expression in lung cells of smokers and nonsmokers. *Am J Resp Cell Mol Biol* 1997, 17, 114-124.
16. Willey JC, Crawford EL, Jackson CM, et al. Expression measurement of many genes simultaneously by quantitative RT-PCR using standardized mixtures of competitive templates. *Am J Resp Cell Mol Biol* 1998, 19, 6-17.
17. Kawasaki H, Shindou K, Higashigawa M, et al. Deoxycytidine kinase mRNA levels in leukemia cells with competitive polymerase chain reaction assay. *Leuk Res* 1996, 20, 677-682.
18. Arnér ESJ, Eriksson S. Mammalian deoxynucleoside kinases. *Pharm Ther* 1995, 67, 155-186.
19. Eriksson S, Kierdaszuk B, Munch-Petersen B, Oberg B, Johansson MG. Comparison of the substrate specificities of human thymidine kinase 1 and 2 and deoxycytidine kinase toward antiviral and cytostatic nucleoside analogs. *Biochem Biophys Res Commun* 1991, 176, 586-592.
20. Kawasaki H, Carrera CJ, Carson DA. Quantitative immunoassay of human deoxycytidine kinase in malignant cells. *Anal Biochem* 1992, 207, 193-196.
21. Brosio O, Spasokoukotskaja T, Arnér ES, et al. Expression of deoxycytidine kinase and phosphorylation of 2-chlorodeoxyadenosine in human normal and tumour cells and tissues. *Eur J Cancer* 1995, 31A, 202-208.
22. Kakiyama T, Fukuda T, Tanaka A, et al. Expression of deoxycytidine kinase (dCK) gene in leukemic cells in childhood: decreased expression of dCK gene in relapsed leukemia. *Leuk Lymphoma* 1998, 31, 405-409.
23. Mansson E, Liliemark E, Soderhall S, Gustafsson G, Eriksson S, Albertioni F. Real-time quantitative PCR assays for deoxycytidine kinase, deoxyguanosine kinase and 5'nucleotidase mRNA measurement in cell lines and in patients with leukemia. *Leukemia* 2002, 31, 386-392.
24. Jacobsson B, Albertioni F, Eriksson S. Deoxynucleoside anabolic enzyme levels in acute myelocytic leukemia and chronic lymphocytic leukemia cells. *Cancer Lett* 2001, 165, 195-200.
25. Arnér E. On the phosphorylation of 2-chlorodeoxyadenosine (CdA) and its correlation with clinical response in leukemia treatment. *Leuk Lymphoma* 1996, 21, 225-231.
26. Colly LP, Peters WG, Richel D, Arentsen-Honders MW, Starrenburg CW, Willemze RH. Deoxycytidine kinase and deoxycytidine deaminase values correspond closely to clinical response to cytosine arabinoside remission induction therapy in patients with acute myelogenous leukemia. *Semin Oncol* 1987, 14(Suppl. 1), 257-261.
27. Stammer G, Zintl F, Sauerbrey A, Volm M. Deoxycytidine kinase mRNA expression in childhood acute lymphoblastoid leukemia. *Anti-Cancer Drugs* 1997, 8, 517-521.
28. Zheng X, Johansson M, Karlsson A. Retroviral transduction of cancer cell lines with the gene encoding drosophila melanogaster multisubstrate deoxyribonucleoside kinase. *J Biol Chem* 2000, 275, 39125-39129.
29. Stegmann AP, Honders WH, Willemze R, Ruiz van Haperen VWT, Landegent JF. Transfection of wildtype deoxycytidine kinase (dCK) cDNA into Ara-C and DAC-resistant rat leukemic cell line of clonal origin fully restores drug sensitivity. *Blood* 1995, 85, 1188-1194.
30. Kaspers GJ, Zwaan CM, Pieters R, Veerman AJ. Cellular drug resistance in childhood acute myeloid leukemia. A mini-review with emphasis on cell culture assays. *Adv Exp Med Biol* 1999, 457, 415-421.
31. Zwaan CM, Kaspers GJ, Pieters R, et al. Different drug sensitivity profiles of acute myeloid and lymphoblastic leukemia and normal peripheral blood mononuclear cells in children with and without Down syndrome. *Blood* 2002, 99, 245-251.
32. Wang LM, Kucera GL. Deoxycytidine kinase is phosphorylated in vitro by protein kinase C alpha. *Biochim Biophys Acta* 1994, 1224, 161-167.
33. Spasokoukotskaja T, Csapo Z, Sasvari-Szekely M, et al. Effect of phosphorylation on deoxycytidine kinase activity. *Adv Exp Med Biol* 2000, 486, 281-285.
34. Csapo Z, Sasvari-Szekely M, Spasokoukotskaja T, Talianidis I, Eriksson S, Staub M. Activation of deoxycytidine kinase by inhibition of DNA synthesis in human lymphocytes. *Biochem Pharmacol* 2001, 61, 191-197.
35. Johansson M, Norda A, Karlsson A. Conserved gene structure and transcription factor sites in the human and mouse deoxycytidine kinase genes. *FEBS Letters* 2000, 487, 209-212.
36. Chen EH, Johnson EE 2nd, Vetter SM, Mitchell BS. Characterization of the deoxycytidine kinase promoter in human lymphoblast cell lines. *J Clin Invest* 1995, 95, 1660-1668.
37. Crawford EL, Peters GJ, Noordhuis P, et al. Standardized RT (StarT)-PCR provides a common language for gene expression. *Mol Diagn* 2001, 6, 217-225.



Expression of deoxycytidine kinase in leukaemic cells compared with solid tumour cell lines, liver metastases and normal liver.

van der Wilt CL, Kroep JR, Loves WJ, Rots MG, Van Groeningen CJ, Kaspers GJ, Peters GJ.

Department of Medical Oncology, VU University Medical Center, Amsterdam, The Netherlands.

Deoxycytidine kinase (dCK) is required for the phosphorylation of several deoxyribonucleoside analogues that are widely employed as chemotherapeutic agents. Examples include cytosine arabinoside (Ara-C) and 2-chlorodeoxyadenosine (CdA) in the treatment of acute myeloid leukaemia (AML) and gemcitabine to treat solid tumours. In this study, expression of dCK mRNA was measured by a competitive template reverse transcriptase polymerase chain reaction (CT RT-PCR) in seven cell lines of different histological origin, 16 childhood and adult AML samples, 10 human liver samples and 11 human liver metastases of colorectal cancer origin. The enzyme activity and protein expression levels of dCK in the cell lines were closely related to the mRNA expression levels ($r=0.75$, $P=0.026$ and $r=0.86$, $P=0.007$). In AML samples, dCK mRNA expression ranged from 1.16 to 35.25 ($\times 10^{-3}$) dCK/beta-actin. In the cell line panel, the range was 2.97-56.9 ($\times 10^{-3}$) dCK/beta-actin of dCK mRNA expression. The enzyme activity in liver metastases was correlated to dCK mRNA expression ($r=0.497$, $P=0.05$). In the liver samples, these were not correlated. dCK mRNA expression showed only a 36-fold range in liver while a 150-fold range was observed in the liver metastases. In addition, dCK activity and mean mRNA levels were 2.5-fold higher in the metastases than in the liver samples. Since dCK is associated with the sensitivity to deoxynucleoside analogues and because of the good correlation between the different dCK measurements in malignant cells and tumours, the CT-RT PCR assay will be useful in the selection of patients that can be treated with deoxycytidine analogues.

PMID: 12628850 [PubMed - indexed for MEDLINE]

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

BBRC

Biochemical and Biophysical Research Communications 327 (2005) 1163–1169

www.elsevier.com/locate/ybbrc

Enhanced expressions of arachidonic acid-sensitive tandem-pore domain potassium channels in rat experimental acute cerebral ischemia

Zheng-Bin Li, Hai-Xia Zhang, Liao-Liao Li, Xiao-Liang Wang*

Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China

Received 11 December 2004

Available online 31 December 2004

Abstract

To further explore the pathophysiological significance of arachidonic acid-sensitive potassium channels, RT-PCR and Western blot analysis were used to investigate the expression changes of TREK channels in cortex and hippocampus in rat experimental acute cerebral ischemia in this study. Results showed that TREK-1 and TRAAK mRNA in cortex, TREK-1 and TREK-2 mRNA in hippocampus showed significant increases 2 h after middle cerebral artery occlusion (MCAO). While the mRNA expression levels of the all three channel subtypes increased significantly 24 h after MCAO in cortex and hippocampus. At the same time, the protein expressions of all the three channel proteins showed significant increase 24 h after MCAO in cortex and hippocampus, but only TREK-1 showed increased expression 2 h after MCAO in cortex and hippocampus. Immunohistochemical experiments verified that all the three channel proteins had higher expression levels in cortical and hippocampal neurons 24 h after MCAO. These results suggested a strong correlation between TREK channels and acute cerebral ischemia. TREK channels might provide a neuroprotective mechanism in the pathological process.

© 2004 Elsevier Inc. All rights reserved.

Keywords: TREK; Potassium channels; Neuroprotection; Cerebral ischemia

The tandem-pore domain potassium channels (K_{2P}) constitute a large subfamily of potassium channel family. Since the identification of its first member in 1996 [1], the channel family has been elucidated further. Altogether 18 K_{2P} genes from the human genome have been cloned, 17 of which have been proved to be functional. According to their amino acid sequence identity and regulation mechanisms, K_{2P} can be divided into six main classes [2]. TREK (the TWIK-related arachidonic acid-stimulated and mechano-gated potassium channels), including TREK-1, TREK-2, and TRAAK, is one class. They have a distinctive structural motif of four transmembrane segments and two pore domains (4TMS/2P) which markedly differs from the structures of classic potassium channel: the 6TMS/1P of voltage-dependent

potassium channels (K_v) and 2TMS/1P of inward-rectifying potassium channels (K_{ir}). In natural and recombinant cells, their currents are simultaneous, voltage- and time-independent, and non-inactivating. Together with other K_{2P} , they produce the baseline or leak current, forming the cellular membrane resting potential [3].

Studies have shown that TREK channels have specific expressions in central nervous system [4]. Electrophysiological experiments showed that they can also be strongly activated by multiple microenvironment parameters in vivo including polyunsaturated fatty acids (PUFAs) such as arachidonic acid (AA), while insensitive to the classic potassium channel antagonists like TEA, 4-AP, and Cs^+ . They are also sensitive to intracellular pH changes and body temperature changes [5,6]. TREK-1 and TREK-2 can be strongly activated by intracellular acidosis, while TRAAK can be activated by intracellular alkalosis [7]. Cell volume, in vitro, can

* Corresponding author. Fax: +86 10 6301 7757.

E-mail address: Wangxl@imm.ac.cn (X.-L. Wang).

also influence their channel activity. The negative pressure applied to the neurons in vitro that mimics the cell volume increase in edema may lead to the channel activation [8]. All the evidence suggests that TREK might play important roles in conditions such as cerebral ischemia and inflammation.

However, the exact functions of TREK channels in such pathological conditions are still not clear until now. In our previous study, it was found that 3 and 30 days after permanent bilateral carotid artery ligation (BCAL), TRAAK mRNA expression increased significantly in rat cortex, while TREK-2 and TRAAK mRNA increased significantly in hippocampus [9]. This indicated that they were involved in chronic cerebral ischemia.

To further elucidate their pathological roles in acute cerebral ischemia, the present study examined the mRNA and protein expressions, by RT-PCR and Western blot analysis, in experimental acute cerebral ischemia in rat cortex and hippocampus. Their histological expression in vivo has also been investigated by immunohistochemistry. By studying their involvement in acute cerebral ischemia in vivo for the first time, this study provided direct evidence for their possible neuro-protective effect in acute cerebral ischemia.

Materials and methods

Experimental acute cerebral ischemia. Adult male Wistar rats weighing 260–300 g were used and were provided by the Animal Center of Chinese Academy of Medical Sciences, Beijing, China. MCAO model was used to produce rat acute cerebral ischemia [10]. Rats were divided into sham, ischemia for 2 h and ischemia for 24 h groups ($n = 5$). Briefly, rats were anesthetized and placed in a supine position. The common carotid (CCA), external carotid (ECA), internal carotid (ICA), and pterygopalatine (PPA) arteries were exposed and ligated except ICA. The cerebral ischemia was produced by inserting a 0.28 mm monofilament suture with a flamed tip about 18–20 mm into ICA from the bifurcation of the ICA and ECA, thus occluding the ostium of the middle cerebral artery (MCA). Sham group rats received the same operation without inserting filament in the MCA. All animals were treated in accordance with the guidelines established by the Experimental Animal Committee of Chinese Academy of Medical Sciences.

Reverse transcription polymerase chain reaction. Analyses of mRNA expressions of TREK-1, TREK-2, and TRAAK in whole areas of rat cortex and hippocampus were performed by reverse transcription polymerase chain reaction (RT-PCR) method according to procedures described previously with modifications [9]. Total cortex and hippocampus RNA was isolated as instructed by the TRIzol Reagent guidelines (Invitrogen, Carlsbad, CA). For cDNA synthesis, 15 μ L reaction mixture containing total RNA 2 μ g and random primers 1 μ g were heated to 70 °C for 5 min. Then dNTP 12.5 mmol, RNasin 25 IU, and M-MLV reverse transcriptase 200 IU (Promega, Madison, WI) were added to a 25 μ L reverse transcription mixture and incubated at 37 °C for 1 h after homogenization. PCR amplification was carried out with 5 μ L cDNA product in a 50 μ L reaction volume containing dNTP 10 mmol, Taq DNA polymerase 2.5 IU (TAKARA Biotechnology, Dalian), and specific oligonucleotide primers 20 pmol. The primer pairs for TREK-1 were 5'-ACAGAACTTCATAGCCCAGCAT-3' and 5'-TCCACCTCTTCCTTCGTCT-3'; those for TREK-2 were 5'-CGTCCATCGTCCAAATT-3' and 5'-TTGGAGGAGTTTCCCTA

CCG-3'; those for TRAAK were 5'-AACTGGTTGCGAGCGGTGTC-3' and 5'-GGGCTTCTTCGTTGGGTGG; and those for β -actin were 5'-CATCTCTTGCTCGAAGTCCA-3' and 5'-ATCATGTTTGAGACCTTCAACA-3' (TAKARA Biotechnology). PCR was performed for 28 or 30 cycles of 94 °C for 1 min, 57 °C for 45 s, and 72 °C for 1 min for β -actin and TREK-1, respectively, and 30 cycles of 94 °C for 1 min, 60 °C for 45 s, and 72 °C for 1 min for TREK-2 and TRAAK. For all the reactions, preliminary experiments were performed to determine the appropriate PCR cycles before they reached the balance stage. Five microliter PCR products were separated by electrophoresis and visualized after ethidium bromide staining under UV light, photographed, and analyzed with Kodak 1D 3.5 system.

Western blot analysis. Levels of protein expressions of the three members of TREK were determined by DAB method as described earlier [11]. Rat cerebral cortex and hippocampus membrane proteins were prepared by being homogenized in lysis buffer (Tris-HCl 5 mM, sucrose 0.32 M, pepstatin A 1 μ g/mL, PMSF 0.2 mM, aprotinin 1 μ g/mL, and leupeptin 1 μ g/mL, pH 7.4) (Sigma-Aldrich, St. Louis, MO); then the lysates were centrifuged at 1000g for 10 min at 4 °C. The supernatants were further subjected to centrifugation at 100,000g for 60 min at 4 °C. The subsequent crude pellet was resuspended in the same solution and protein concentration was detected by the Bradford method. All procedures were carried out on ice to prevent proteolysis.

For electrophoresis, the equivalents of 100 μ g protein were applied to each lane of a 10% SDS-PAGE gel and run at standard conditions. After that, the proteins were electroblotted onto nitrocellulose membrane (0.45 μ m, Amersham, Buckinghamshire) at 1 mA/cm² for 2.5 h by semi-dry apparatus (Liuyi, Beijing). Membranes were blocked in 5% non-fat dried milk in TBST (100 mM Tris-HCl, 154 mM NaCl, and 0.05% Tween 20 (v/v), pH 7.5) for 2 h at room temperature with agitation, followed by being incubated with polyclonal primary antibodies of TREK-1, TREK-2, and TRAAK, and β -actin (1:1000 dilution with TBST, Santa Cruz Biotechnology, CA) at 4 °C overnight. After that, the membranes were incubated with secondary antibodies (rabbit anti-goat IgG conjugated to horseradish peroxidase, HRP, 1:5000 dilution, Santa Cruz) at room temperature for 2 h. The antigen-antibody complexes were visualized with 3,3'-diaminobenzidine-tetrahydrochloride (DAB, Research Genetics, Huntsville, AL), and photographed and analyzed with Kodak 1D 3.5 system. All data were expressed as ratios to β -actin protein. The specificity of primary antibodies was tested by pre-absorption with the antigenic peptide.

Immunohistochemistry. Immunohistochemistry was performed on cryosections as described previously [12]. Rats of sham group and 24 h after MCAO group were transcardially perfused with 4% paraformaldehyde after being thoroughly anesthetized. Next, the brains were removed, post-fixed in 4% paraformaldehyde at 4 °C for 4 h, and transferred to 30% sucrose and 4% paraformaldehyde solution at 4 °C for 48 h until the brain sank to the bottom. After being frozen in OCT compound (Jung, Leica, Germany) for 10 s in liquid nitrogen, the brains were cut into 40 μ m coronal sections in a cryostat (CM1900, Leica). Sections were consecutively kept in 3% hydrogen peroxide for 15 min to quench the endogenous peroxidase activity and 5% rabbit serum for 30 min to block. They were then incubated with primary antibodies (1:50 dilution, Santa Cruz Biotechnology) for 2 h at 37 °C and washed in PBST three times for 10 min. Following this, sections were incubated with HRP-conjugated secondary antibodies (1:100 dilution, rabbit anti-goat IgG, Santa Cruz Biotechnology) at room temperature for 2 h with gentle agitation. Then sections were visualized with DAB and slide-mounted, dried, defatted, and coverslipped. The specificity of the antibodies was assessed by pre-absorption with the antigenic peptide. For quantification, five positively stained neurons were randomly selected and analyzed with Image Pro Plus software.

Statistical analysis. All the data of RT-PCR, Western blot analysis, and immunohistochemical results were analyzed with one-way ANOVA tests. Results were expressed as means \pm SD. And significance was considered at $p < 0.05$.

Results

mRNA expression changes of TREK-1, TREK-2, and TRAAK in cortex and hippocampus after MCAO

mRNA expression changes of TREK-1, TREK-2, and TRAAK were detected in both cortex and hippocampus 2 and 24 h after MCAO. The specificity of RT-PCR products was confirmed by the results of agarose gel electrophoresis that yielded single product bands of the expected sizes in all cases. Fig. 1 shows that in cortex, 2 h after MCAO, TREK-1 and TRAAK mRNA expressions increased significantly, while all the three channel mRNA showed significantly enhanced expressions 24 h after MCAO compared with the sham group. In hippocampus, TREK-1 and TREK-2 mRNA expressions increased significantly both 2 and 24 h after MCAO. But TRAAK mRNA showed enhanced expression only 24 h after MCAO. Results also showed that the mRNA expression levels of the three channels in hippocampus are lower than those in cortex.

Protein expression changes of TREK-1, TREK-2, and TRAAK in cortex and hippocampus after MCAO

To detect the expression changes of TREK-1, TREK-2, and TRAAK, whole cortex and hippocampus were

subjected to the isolation of the membrane proteins. So the expression changes of related channel proteins represented their expressions of the whole areas. Expressions of the three channel proteins were measured by Western blot analysis 2 and 24 h after MCAO, respectively. Fig. 2 showed that 24 h after MCAO, all the three channel proteins showed significantly increased expressions in cortex, indicating that more functional channel proteins have been synthesized. But only TREK-1 showed higher expression 2 h after MCAO, while TREK-2 and TRAAK did not show statistically significant changes due to the supposed reason of time delay between the gene activation and protein expression. At the same time, all the three proteins also showed significantly increased expressions in hippocampus 24 h after MCAO, while 2 h after MCAO, only TREK-1 channel protein showed significant increases in hippocampus.

Histological expressions of TREK-1, TREK-2, and TRAAK channel proteins in cortical and hippocampal neurons 24 h after MCAO

In the study, the expression changes of TREK-1, TREK-2, and TRAAK in PCg region of cortex and CA1 region of hippocampus were analyzed because these two regions are most vulnerable to ischemic insults in CNS. Immunohistochemical experiments showed that

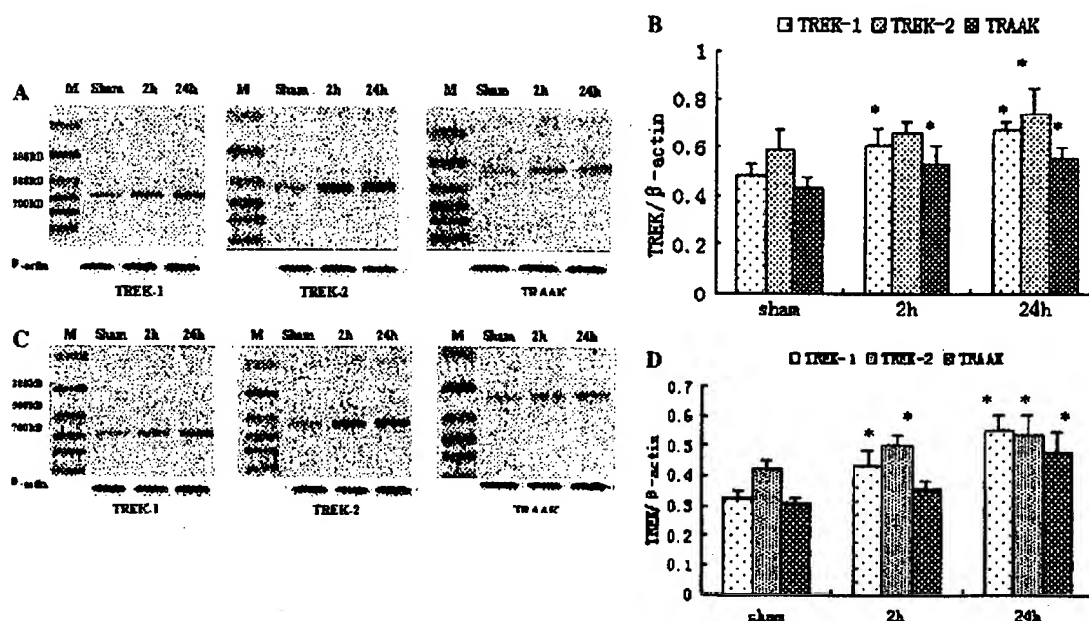


Fig. 1. Expression changes of TREK-1, TREK-2, and TRAAK mRNA in cortex and hippocampus 2 and 24 h after MCAO. Sham, 2, and 24 h represent sham group, 2, and 24 h after MCAO, respectively. (A) Representative ethidium bromide-stained agarose gel of RT-PCR products of TREK-1, TREK-2, TRAAK, and β -actin of cortex after electrophoresis. Five microliter RT-PCR products were loaded to each lane, respectively. (B) Quantification of the relative expression levels of TREK-1, TREK-2, and TRAAK in cortex after acute MCAO. (C) Representative ethidium bromide-stained agarose gel of RT-PCR products of TREK-1, TREK-2, TRAAK, and β -actin of hippocampus after electrophoresis. (D) Quantification of the relative expression levels of TREK-1, TREK-2, and TRAAK in hippocampus after MCAO. Data were acquired by image analysis of the RT-PCR bands and were ratios compared with β -actin. Each value represents the mean densitometric value of three repeated RT-PCRs. $n = 5$ for each group. * $p < 0.01$ compared with the sham group.

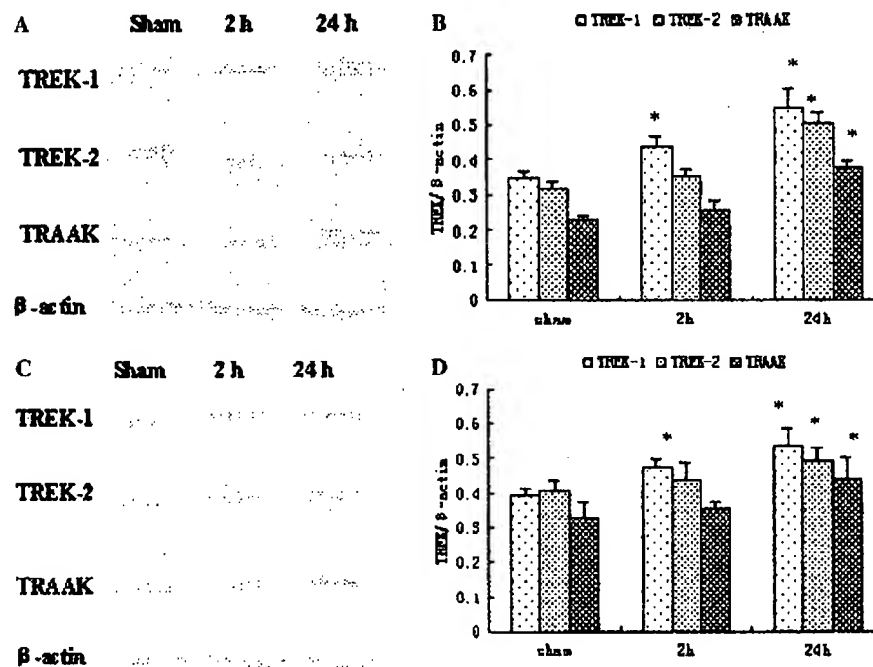


Fig. 2. Expression changes of TREK-1, TREK-2, and TRAAK channel proteins in cortex and hippocampus 2 and 24 h after MCAO. Sham, 2, and 24 h represent sham group, 2, and 24 h after MCAO, respectively. (A) Representative Western blots of TREK-1, TREK-2, TRAAK, and β -actin proteins detected by DAB staining in cortex. (B) Quantification of Western blot signals of TREK-1, TREK-2, and TRAAK in cortex after MCAO. (C) Representative Western blots of TREK-1, TREK-2, TRAAK, and β -actin proteins detected by DAB staining in hippocampus. (D) Quantification of Western blot signals of TREK-1, TREK-2, and TRAAK in hippocampus. All data were ratios compared with β -actin protein. Each value represents the mean of three membrane preparations normalized to the densitometric value. Values are means \pm SD. $n = 5$ for each group. * $p < 0.01$ compared with the control group.

in both these two regions, less positively stained neurons were detected 24 h after MCAO compared to the sham group, because many have died following MCAO surgery (Fig. 3). But for the remaining neurons in both PCg and CA1 regions, significantly increased signals of TREK-1, TREK-2, and TRAAK proteins were detected 24 h after MCAO as shown in Table 1.

Discussion

Stroke, including occlusive stroke, is one of the leading causes of death and dependence in activities of daily living in the world. It can result in severe neuronal degeneration and consequent loss of brain functions. Recent developments have further advanced our knowledge of the cellular and molecular mechanisms, and have confirmed that cerebral ischemia gives rise to widespread alteration of gene expressions and regulations. However, the pathophysiological process and the protective mechanisms following brain injury induced by ischemia are extremely complex and have not been fully elucidated.

TREK-1, TREK-2, and TRAAK belong to TREK class of the K_{2P} family. Previous studies have proved

that they are highly expressed in regions prone to ischemia-induced cell death such as cortex, hippocampus, and amygdala, but expressed in much lower level in regions like spinal cord, brainstem, and substantia nigra that are evolutionarily more ancient [4]. While animal models for transient forebrain ischemia identified that cortex and hippocampus are very vulnerable to ischemic stroke, especially hippocampus. And within the hippocampus, pyramidal cells in the CA1 region are most susceptible to ischemic insults. TREK-1, TREK-2, and TRAAK can be strongly activated by pathological microenvironmental changes such as membrane distortion, intracellular pH increase/decrease [13]. Since these pathological changes usually occur in cerebral ischemia, the above-mentioned channel subtypes are suggested to be involved in the pathological process.

When cerebral ischemia occurs, the activation of phospholipase A_2 , induced by intracellular Ca^{2+} release, resulted in degradation of membrane phospholipids and the accumulation of unsaturated fatty acids like AA, as well as lysophosphatidylcholine (LPC) and lysophosphatidic acid (LPA) [14]. At the same time, the non-oxygen and low-glucose condition in brain can lead to the increase of glycolysis and consequent lactate concentration increase, resulting in intracellular acidosis. Also, the

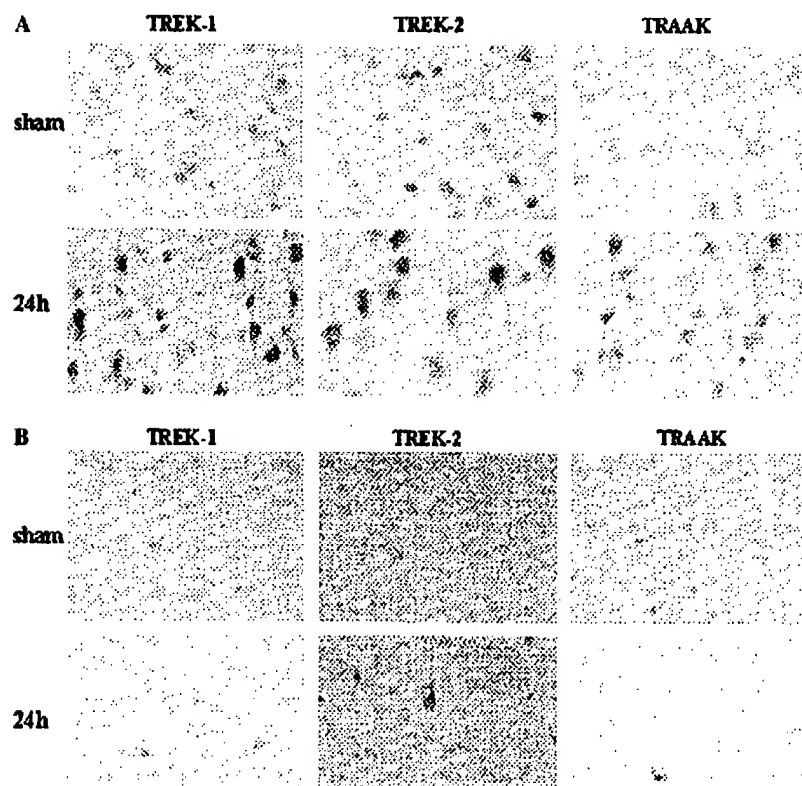


Fig. 3. Immunohistochemical analysis of TREK-1, TREK-2, and TRAAK channel protein expressions in cortical and hippocampal neurons 24 h after MCAO. Sham and 24 h represent sham group and 24 h after MCAO, respectively. (A) Positive neurons detected by DAB staining in cortical cryosections of PCg area. (B) Positive neurons detected by DAB staining in hippocampal cryosections of CA1 area. Scale bar = 0.01 mm.

Table 1

Enhanced expressions of TREK channel proteins in cortical and hippocampal neurons 24 h after MCAO

Channel protein	Sham	24 h
(A)		
TREK-1	83.99 ± 7.33	63.95 ± 1.83*
TREK-2	75.12 ± 5.28	55.59 ± 2.92*
TRAAK	92.32 ± 5.60	70.07 ± 1.05*
(B)		
TREK-1	123.67 ± 5.30	89.46 ± 1.56*
TREK-2	116.21 ± 4.37	79.26 ± 3.43*
TRAAK	102.63 ± 4.18	81.67 ± 2.02*

Sham and 24 h represent sham group and 24 h after MCAO, respectively. Five positively stained neurons were randomly selected and analyzed in each cortical or hippocampal section. Each value represents the mean of five positively stained neurons normalized to the densitometric value. (A) Analysis results of cortical sections. (B) Analysis results of hippocampal sections. Values are means ± SD. $n = 3$ for each group.

* $p < 0.01$ compared with the control group.

edema caused in ischemia mimics the function of negative mechanical stretch. All these factors are potent activators of TREK-1, TREK-2, and TRAAK in the case of acute cerebral ischemia (intracellular acidosis can activate TREK-1 and TREK-2, but not TRAAK). In our

experiments, TREK-1, TREK-2, and TRAAK showed enhanced expressions both in cortex and hippocampus at mRNA and protein levels in experimental acute cerebral ischemia. Although many cells died, the remaining neurons had markedly higher expression levels of TREK channels 24 h after MCAO surgery, indicating that more functional proteins have been synthesized in the living neurons. This might be the reason why they survive. So, it seems that in the short period after cerebral ischemia, the TREK channels were activated by the pathological changes in the microenvironment, but in a relatively long period, their genes were activated and more functional proteins were expressed.

In physiological conditions, TREK-1, TREK-2, and TRAAK can produce leak current that determines the resting potential, and thus influences the cell excitability. Usually, when potassium channels are activated, the efflux of K^+ from cells can lead to membrane hyperpolarization, which can decrease the cell excitability. In the case of ischemia, the decrease of cell excitability has protective effect by reducing Ca^{2+} influx and release of stimulative neurotransmitters. This is the same protective mechanism as ATP-sensitive potassium channels (K_{ATP}) which can be activated in the same condition and acts as a protective agent by counterbalancing the

depolarization induced by glutamine, 5-hydroxytryptamine (5-HT), etc. [15]. So, the activation of TREK genes and enhanced expression of functional channel proteins following MCAO can counteract the excitotoxicity induced by stimulative neurotransmitters such as glutamate and 5-HT. In this way, it seems that the activation of TREK-1, TREK-2, and TRAAK in cerebral ischemia provides a neuroprotective effect. This was supported by other studies. These studies showed that neuroprotective agents such as riluzole and some volatile anesthetics can also activate TREK-1, TREK-2, and TRAAK [16,17].

It was also reported that TREK-2 can also be regulated by Gs and Gi that were linked to 5-HT and glutamate receptors [18]. This indicates that its activity is closely related to these excitatory neurotransmitters and may be in charge of tuning neuronal excitability in response of a variety of neurotransmitters and hormones.

But some scholars hold different opinions. In one study, sipatrigine as a neuroprotective agent showed the function of potent inhibition of TREK-1, TREK-2, and TRAAK activity. But actually, no evidence showed that sipatrigine provides protective effect by inhibiting TREK-1, TREK-2, and TRAAK, and this might only be a side-effect [19]. And recently, P_{O_2} or O_2 concentration has been supposed to produce inhibitory effect to TREK-1 activity, thus casting a doubt on the role TREK-1 may play in neuroprotection in cerebral ischemia [20,21]. But because of the lack of specific antagonists, there is still much work to do to fully elucidate the question.

Cerebral ischemia is a complex pathological process. There are still uncertainties as to the protective mechanisms and it may involve a cascade of pathological molecular events. Our study once again provided direct evidence for the correlation between TREK channels and cerebral ischemia. TREK-1, TREK-2, and TRAAK might provide an important neuroprotective mechanism in the pathological process following acute cerebral ischemia.

Acknowledgments

This study was supported by National 863 Foundation Project of China, No. 2004AA223815. We thank Prof. Lin Li for her help in the immunohistochemical experiments. We are also grateful to Dr. Gareth Davey for his help in this research.

References

- [1] F. Lesage, E. Guillemarc, M. Fink, F. Duprat, M. Lazdunski, G. Romey, J. Barhanin, TWIK-1, a ubiquitous human weakly inward rectifying K^+ channel with a novel structure, *EMBO J.* 15 (5) (1996) 1004-1011.
- [2] A.J. Patel, F. Honore, Properties and modulation of mammalian 2P domain K^+ channels, *Trends Neurosci.* 24 (6) (2001) 339-346.
- [3] F. Maingret, E. Honore, M. Lazdunski, A.J. Patel, Molecular basis of the voltage-dependent gating of TREK-1, a mechanosensitive $K(+)$ channel, *Biochem. Biophys. Res. Commun.* 292 (2) (2002) 339-346.
- [4] E.M. Talley, G. Solorzano, Q. Lei, D. Kim, D.A. Bayliss, CNS distribution of members of the two-pore-domain (KCNK) potassium channel family, *J. Neurosci.* 21 (19) (2001) 7491-7505.
- [5] H. Bang, Y. Kim, D. Kim, TREK-2, a new member of the mechanosensitive tandem-pore K^+ channel family, *J. Biol. Chem.* 275 (23) (2000) 17412-17419.
- [6] D. Kim, Fatty acid-sensitive two-pore domain K^+ channels, *Trends Pharmacol. Sci.* 24 (12) (2003) 648-654.
- [7] F. Maingret, A.J. Patel, F. Lesage, M. Lazdunski, E. Honore, Lysophospholipids open the two-pore domain mechano-gated $K(+)$ channels TREK-1 and TRAAK, *J. Biol. Chem.* 275 (14) (2000) 10128-10133.
- [8] C. Heurteaux, N. Guy, Laigle, C.N. Blondeau, F. Duprat, M. Mazzuca, L. Lang-Lazdunski, C. Widmann, M. Zanzouri, G. Romey, M. Lazdunski, TREK-1, a $K(+)$ channel involved in neuroprotection and general anesthesia, *EMBO J.* (2004).
- [9] X.H. Xu, Y.P. Pan, X.L. Wang, Alterations in the expression of lipid and mechano-gated two-pore-domain potassium channel genes in rat brain following chronic cerebral ischemia, *Brain Res. Mol. Brain Res.* 120 (2) (2004) 205-209.
- [10] C.M. Maier, K. Abern, M.L. Cheng, J.E. Lee, M.A. Yenari, G.K. Steinberg, Optimal depth and duration of mild hypothermia in a focal model of transient cerebral ischemia: effects on neurologic outcome, infarct size, apoptosis, and inflammation, *Stroke* 29 (10) (1998) 2171-2180.
- [11] A.S. Alberts, M. Montminy, S. Shenolikar, J.R. Feramisco, Expression of a peptide inhibitor of protein phosphatase 1 increases phosphorylation and activity of CREB in NIH 3T3 fibroblasts, *Mol. Cell. Biol.* 14 (7) (1994) 4398-4407.
- [12] S.N. Dixit, J.M. Seyer, A.H. Kang, Biochemical and immunohistochemical characterization and internal alignment of pepsin-derived collagenous fragments of the alpha 1(IV) chain from bovine kidney cortices, *J. Biol. Chem.* 257 (9) (1982) 4864-4868.
- [13] I. Lauritzen, N. Blondeau, C. Heurteaux, C. Widmann, G. Romey, M. Lazdunski, Polyunsaturated fatty acids are potent neuroprotectors, *EMBO J.* 19 (8) (2000) 1784-1793.
- [14] N.G. Wahlgren, N. Ahmed, Neuroprotection in cerebral ischemia: facts and fancies—the need for new approaches, *Cerebrovasc. Dis.* 17 (Suppl. 1) (2004) 153-166.
- [15] D. Enkvetchakul, C.G. Nichols, Gating mechanism of K_{ATP} channels: function fits form, *J. Gen. Physiol.* 122 (5) (2003) 471-480.
- [16] A.J. Patel, E. Honore, F. Lesage, M. Fink, G. Romey, M. Lazdunski, Inhalational anesthetics activate two-pore-domain background K^+ channels, *Nat. Neurosci.* 2 (5) (1999) 422-426.
- [17] F. Duprat, F. Lesage, A.J. Patel, M. Fink, G. Romey, M. Lazdunski, The neuroprotective agent riluzole activates the two P domain $K(+)$ channels TREK-1 and TRAAK, *Mol. Pharmacol.* 57 (5) (2000) 906-912.
- [18] F. Lesage, C. Terrenoire, G. Romey, M. Lazdunski, Human TREK-2, a 2P domain mechano-sensitive K^+ channel with multiple regulations by polyunsaturated fatty acids, lysophospholipids, and Gs, Gi, and Gq protein-coupled receptors, *J. Biol. Chem.* 275 (37) (2000) 28398-28405.
- [19] H.J. Meadows, C.G. Chapman, D.M. Duckworth, R.E. Kelsell, P.R. Murdock, S. Nasir, G. Rennie, A.D. Randall, The neuroprotective agent sipatrigine (BW619C89) potently inhibits the human tandem pore-domain $K(+)$ channels TREK-1 and TRAAK, *Brain Res.* 892 (1) (2001) 94-101.

[1] F. Lesage, E. Guillemarc, M. Fink, F. Duprat, M. Lazdunski, G. Romey, J. Barhanin, TWIK-1, a ubiquitous human weakly

Z.-B. Li et al. / *Biochemical and Biophysical Research Communications* 327 (2005) 1163-1169

1169

[20] P.J. Kemp, C. Peers, A. Lewis, P. Miller, Regulation of recombinant human brain tandem P domain K^+ channels by hypoxia: a role for O_2 in the control of neuronal excitability? *J. Cell Mol. Med.* 8 (1) (2004) 38-44.

[21] P. Miller, C. Peers, P.J. Kemp, Polymodal regulation of hTREK1 by pH, arachidonic acid, and hypoxia: physiological impact in acidosis and alkalosis, *Am. J. Physiol. Cell Physiol.* 286 (2) (2004) C272-C282.

ORIGINAL ARTICLE

Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics

GM Nagaraja¹, M Othman², BP Fox¹, R Alsaber¹, CM Pellegrino³, Y Zeng², R Khanna², P Tamburini³, A Swaroop² and RP Kandpal¹¹Department of Biological Sciences, Fordham University, Bronx, NY, USA; ²Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI, USA and ³Bayer Corporation, West Haven, CT, USA

We have characterized comprehensive transcript and proteomic profiles of cell lines corresponding to normal breast (MCF10A), noninvasive breast cancer (MCF7) and invasive breast cancer (MDA-MB-231). The transcript profiles were first analysed by a modified protocol for representational difference analysis (RDA) of cDNAs between MCF7 and MDA-MB-231 cells. The majority of genes identified by RDA showed nearly complete concordance with microarray results, and also led to the identification of some differentially expressed genes such as lysyl oxidase, copper transporter ATP7A, EphB6, RUNX2 and a variant of RUNX2. The altered transcripts identified by microarray analysis were involved in cell–cell or cell–matrix interaction, Rho signaling, calcium homeostasis and copper-binding/sensitive activities. A set of nine genes that included GPCR11, cadherin 11, annexin A1, vimentin, lactate dehydrogenase B (upregulated in MDA-MB-231) and GREB1, S100A8, amyloid β precursor protein, claudin 3 and cadherin 1 (downregulated in MDA-MB-231) were sufficient to distinguish MDA-MB-231 from MCF7 cells. The downregulation of a set of transcripts for proteins involved in cell–cell interaction indicated these transcripts as potential markers for invasiveness that can be detected by methylation-specific PCR. The proteomic profiles indicated altered abundance of fewer proteins as compared to transcript profiles. Antisense knockdown of selected transcripts led to inhibition of cell proliferation that was accompanied by altered proteomic profiles. The proteomic profiles of antisense transfectants suggest the involvement of peptidyl-prolyl isomerase, Raf kinase inhibitor and 80 kDa protein kinase C substrate in mediating the inhibition of cell proliferation.

Oncogene (2006) 25, 2328–2338. doi:10.1038/sj.onc.1209265; published online 28 November 2005

Keywords: representational difference analysis; microarrays; proteomics; breast carcinoma; biomarkers; copper homeostasis

Introduction

The transformation of a normal cell into a cancer cell has been correlated to altered expression of a variety of genes (Perou *et al.*, 2000; Becker *et al.*, 2005). The expression of some of these genes is a direct result of sequence mutation, whereas other changes occur due to alterations in gene products that participate in specific pathways. The changes in gene expression have been routinely characterized by classical subtraction hybridization and differential display approaches (Cerosaletti *et al.*, 1995; Alpan *et al.*, 1996). With the availability of the human genome sequence and sequences for a number of other model organisms, traditional methods have largely been replaced by gene microarrays (Khan *et al.*, 2001). These analyses have been used to characterize the molecular basis of a variety of diseases including cancer. A comprehensive analysis of a large number of cancer cell lines allowed clustering of genes into groups based on their expression patterns in phenotypically related cell lines (Khan *et al.*, 2001; Dan *et al.*, 2002; Rosenwald *et al.*, 2002; van't Veer *et al.*, 2002). The results of profiling experiments indicated expression of specific gene clusters in cell lines that have the same origin or have arisen from the same organ (Ross *et al.*, 2000). A complementary approach that has been used in limited ways is proteomics. Proteomics scores for changes in different proteins and peptides in cells with characteristic phenotypic differences. However, a comparative analysis of transcripts and proteins to establish a relationship between transcript changes and protein levels has not yet become routine.

Although expression profiling of tumor tissue and its comparison with normal tissue, in principle, is most appropriate to obtain the genetic signatures of a tumor type, such comparisons have not been free of attendant

Correspondence: Dr RP Kandpal, Department of Biological Sciences, Larkin Hall, Room 250, 441 E Fordham Road, Fordham University, Bronx, NY 10458, USA.

E-mail: kandpal@fordham.edu

Received 29 September 2005; revised 14 October 2005; accepted 14 October 2005; published online 28 November 2005

complications. These complications arise due to heterogeneity of tumor specimens wherein any cell type-specific changes are likely to be masked by other cell types that constitute the tumor specimen. For this reason, well-characterized cell lines established from tumor tissue may prove more informative and have been considered useful by cancer researchers. Comparing gene profiles between cell lines has the potential to reveal genes that could be causative for the phenotype and other genes that can serve as tumor biomarkers.

Our investigations are aimed at designating a subset of transcripts that could distinguish a normal breast cell from a breast cancer cell and help to predict tumorigenic or metastatic potential of a transformed cell. We describe here transcript and proteomic profiles of a normal breast cell line, a tumorigenic but noninvasive breast carcinoma cell line and an invasive breast carcinoma cell line, and summarize them as a set of candidate biomarkers or targets for therapeutic intervention. The comparison of transcript profiles with proteomic profiles demonstrated that altered proteins were not always represented in the microarray designated profiles and *vice versa*. Furthermore, we have targeted five transcripts that were upregulated in MCF7 cells for investigating their role in cell proliferation pathways. The proteomic profiles have revealed that inhibition of cell proliferation by antisense knockdown was mediated by a specific set of proteins.

Results

Representational difference analysis

As described in the Materials and methods section, RDA was performed by using cDNAs from MCF7 and MDA-MB-231 as tester/driver or driver/tester combination. The difference product in the first case represents the genes that are either upregulated in or specific to MCF7. On the other hand, the difference product of MDA-MB-231 (tester) and MCF7 (driver) hybridization resulted in the isolation of cDNAs that are either upregulated in or specific to MDA-MB-231. The initial linkers used in this protocol had internal *Bgl*II sites. One strand of the linker was used to amplify both the tester and driver cDNAs after linkers had been ligated to cDNAs. After removal of linkers from amplified cDNAs by digestion with *Bgl*II, a dephosphorylated *Bgl*II adaptor was ligated to tester DNA. The *Bgl*II adaptor had an internal *Eco*RI site. The difference product was digested with *Eco*RI and cloned in pBlueScript vector. The cloning efficiency of the difference product was very low (5×10^4 c.f.u./ μ g of DNA). The low efficiency of cloning is attributed to a substantial fraction of amplified DNA product that is refractory to restriction digestion. The sequencing of a set of 100 clones each from the difference libraries revealed 50 different kinds of clones. The majority of these sequences were short fragments and represented either 3' regions or internal fragments of transcripts. A summary of these clones is presented in Table 1. The

involvement of the majority of these cDNAs is well characterized either in tumorigenesis or in metastasis. The phenotypic characteristics of MCF-7 and MDA-MB-231 ideally match with the biological significance of these genes. The alterations in transcripts for Rho signaling proteins, Ca^{2+} binding/requiring proteins, tight junctions/anchoring junctions/gap junctions, copper binding or sensitive proteins, and RUNX2 are particularly noteworthy.

The differential expression of a representative number of RDA clones was validated by semiquantitative PCR. As shown in Figure 1, these transcripts were either specific to or upregulated in the cell line that was used as a tester. Such analyses demonstrated that more than 90% of the clones were differentially expressed. The abundance of transcripts and the results of RT-PCR were also confirmed by Northern blotting (Figure 2). The pattern of hybridization clearly indicates that all these transcripts showed differential expression in MCF7 and MDA-MB-231 cells that were used as driver/tester combinations for the RDA.

Gene microarrays

After obtaining preliminary molecular signatures of MCF7 and MDA-MB-231 cells by RDA, we used Affymetrix gene microarrays to expand the above analysis to identify a comprehensive set of transcripts that is deregulated in invasive breast carcinoma cells. The comparisons of cell lines on the basis of transcripts that are either present or absent as shown in Figure 3 revealed that a set of 123 genes distinguishes MDA-MB-231 cells from MCF7 and MCF10A. These genes can be classified by their involvement in functional classes such as transcription, signal transduction, cell adhesion, cell cycle, metabolism, transport, response genes and development (Figure 4). The majority of these genes participated in the process of signal transduction followed by transcription, cell adhesion and metabolism, respectively. A few transcripts in these classes were tested by real-time RT-PCR to confirm their altered abundance. The selected transcripts showed changes ranging between two- and 10-fold, 11- and 20-fold and greater than 20-fold, and were in close agreement with the results of microarray analysis. The qualitative pattern of change observed in microarrays analysis was readily reproduced by real-time or semiquantitative RT-PCR for all transcripts tested.

The number of altered transcripts was over 1000 based on a change of twofold or greater, and a majority of these genes show changes varying between two- and fourfold (Figure 5). Interestingly, with all comparisons combined, there were 21 genes downregulated more than 50-fold and 55 genes that were upregulated more than 18-fold when specific cell line pairs were compared (Figure 5). The transcripts that represent the extremes of upregulated and downregulated scale can allow distinction between MCF7 and MDA-MB-231 cells. These transcripts include GPCR11, cadherin 11, annexin A1, vimentin, lactate dehydrogenase B (upregulated in MDA-MB-231) and GREB1, S100A8, amyloid

Table 1 Differentially expressed transcripts identified by RDA

Upregulated in MDA-MB-231	Downregulated in MDA-MB-231
Extracellular matrix/matrix-crosslinking proteins LOX Laminin β 1 Collagen VI- α 1	Calcium-binding proteins Calgranulin B
Calcium-binding proteins Reticulocalbin 1 S100A8 Cullin 5	Transcription factors/promoter-binding proteins: Chromosome 4 ORF Estrogen receptor 1 RUNX2 variant (exon 8 deleted)
Transcription factors/promoter-binding proteins RUNX2 c-Jun Fra-1	Cell-cell adhesion/cell-surface receptor proteins Claudin 3 Amyloid β precursor protein Triose phosphate isomerase Plakoglobin Cdh 1 Cdh 3 Annexin A9 RAR- α Connexin 31
Cell-cell adhesion/cell-surface receptor proteins: Cdh11 CYR61 MHC class II antigen γ chain Protease-activated receptor-1 Protease-activated receptor-4	ATPase/GTPase and signal transduction proteins RhoD RhoB TGF- β R1
ATPase/GTPase and signal transduction proteins ATP7a Caveolin 2 AXL receptor tyrosine kinase Rho GEF 3 Rho/Rac GEF 18 P21-Rac2	Stress-response proteins Protein kinase H11
Metalloproteases and MMP inhibitor proteins TIMP-2 MT1-MMP	Cytoskeletal component and binding proteins Keratin 18 Tubulin δ 1 Microtubule-associated protein τ
Stress-response proteins Dual specificity phosphatase (DUSP)	Cell-cycle regulation and growth/differentiation/apoptosis proteins S100A13 S100C Aurora kinase AIK2 Nucleosidediphosphate kinase
Cytoskeletal component and binding proteins Moesin Vimentin Filamin B	Secreted proteins and growth factors Trefol factor 3 (TFF3) Trefol factor 1 (TFF1) Four and a half LIM domain1 Solute carrier family 16 SLC16A6
Cell-cycle regulation and growth/differentiation/apoptosis proteins Cyclin B1 Cyclin E Cyclin A2 Bcl2-like 1 protein	DNA replication DNA replication complex GINS-PSF2
Secreted proteins and growth factors Milk fat globule protein TGF- α SMAD-specific E3 ub ligase 2	Miscellaneous Serine protease inhibitor type 1 (SPINT1) Human homolog of <i>Xenopus</i> protein XAG Hypothetical protein FLJ22222 Hypothetical protein 20171 Hypothetical protein MGC3265
Miscellaneous Prion protein	

β precursor protein, claudin 3 and cadherin 1 (downregulated in MDA-MB-231). The distinction between MCF7 and MCF10A may be made based on keratin 19, serine protease, amyloid β precursor, neuropeptide Y receptor Y1 (upregulated in MCF7) and caldesmon, annexin A1, epithelial membrane protein 1, S100A2,

keratin 15 (downregulated in MCF7). Likewise, MDA-MB-231 cells differ from MCF10A in vimentin, epithelial membrane protein 3, cadherin 11, GPCR 116, collagen type XIII α 1, Bcl2-associated athanogene 2 (upregulated in MDA-MB-231) and keratin 15, cystatin A, cadherin 1, CD24, calcium-activated chlor-

ide channel, S100P, GPCR 87 (downregulated in MDA-MB-231). Thus, a small subset of transcripts may serve as an accurate signature of these cell lines. Several of these gene products have been shown to participate in tumorigenesis and invasiveness of breast carcinoma cell lines.

The invasiveness phenotype of MDA-MB-231 cells specifically relates to changes in the following functional classes: (a) cell adhesion molecules, (b) Ca^{2+} requiring, Ca^{2+} binding or Ca^{2+} regulatory genes, (c) copper-sensitive or copper-transporting proteins and (d) specific regulatory proteins of Rho signaling. Among these functional groups, 23 transcripts involved in cell-cell or cell-matrix interactions are underexpressed in MDA-

MB-231 cells and 21 transcripts were overexpressed in this cell line (Table 2). The comparison of Ca^{2+} -requiring/binding genes indicated downregulation of 26 transcripts and upregulation of 26 transcripts in invasive cells as compared to noninvasive cells (Table 3). While Ca^{2+} homeostasis is extensively investigated in human cancers, copper homeostasis is an underexplored area. The alterations in copper homeostasis in breast carcinoma cells were reflected by changes in transcripts corresponding to a variety of copper-binding or copper-sensitive proteins/enzymes (Tables 1 and 4). The deregulation of Rho signaling was evident from changes in various proteins involved in this pathway (Table 4).

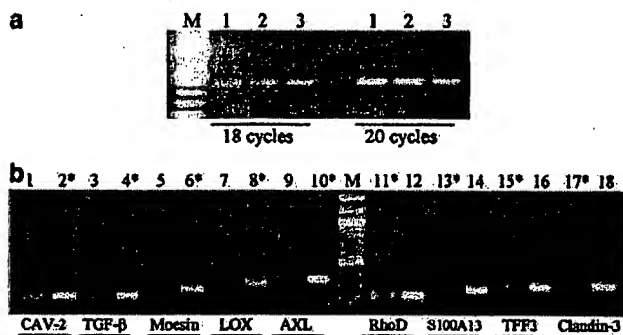


Figure 1 Semiquantitative evaluation of selected transcripts. (a) RNA was isolated from confluent culture dishes containing MCF10A (lanes 1 and 4), MCF7 (lanes 2 and 5) or MDA-MB-231 (lanes 3 and 6) cells. The amount of RNA was first determined spectrophotometrically. Equal amounts of RNA, as determined by absorbance at 260 nm, were amplified with primers specific to actin gene for 18 cycles (lanes 1–3) or 20 cycles (lanes 4–6). The lane containing size markers is labeled as M. (b) A set of primers corresponding to caveolin 2 (lanes 1 and 2), TGF- α (lanes 3 and 4), Moesin (lanes 5 and 6), LOX (lanes 7 and 8), Axl receptor (lanes 9 and 10), RhoD (lanes 11 and 12), S100A13 (lanes 13 and 14), TFF3 (lanes 15 and 16) and Claudin 3 (lanes 17 and 18) were amplified for 32 cycles. Lanes 1, 3, 5, 7, 9, 12, 14, 16 and 18 represent amplified products from MCF7 and lanes 2, 4, 6, 8, 10, 11, 13, 15 and 17 represent MDA-MB-231 cells. The lanes containing PCR products from MDA-MB-231 cells are marked with an asterisk.

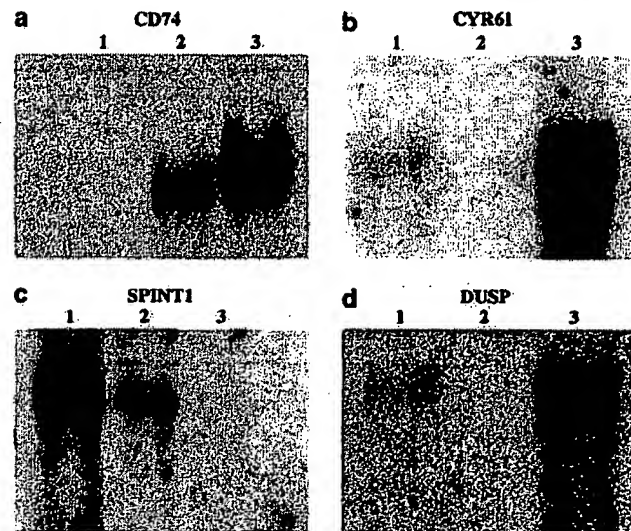


Figure 2 Analysis of selected transcripts by Northern hybridization. The probes specific to CD74 (a), CYR61 (b), SPINT1 (c) and DUSP (d) were labeled with a ^{32}P nucleotide and hybridized to blots containing RNA from MCF10A (lane 1), MCF7 (lane 2) and MDA-MB-231 (lane 3). The blots were washed stringently and developed as described. The amounts of RNA loaded were normalized as in Figure 1.

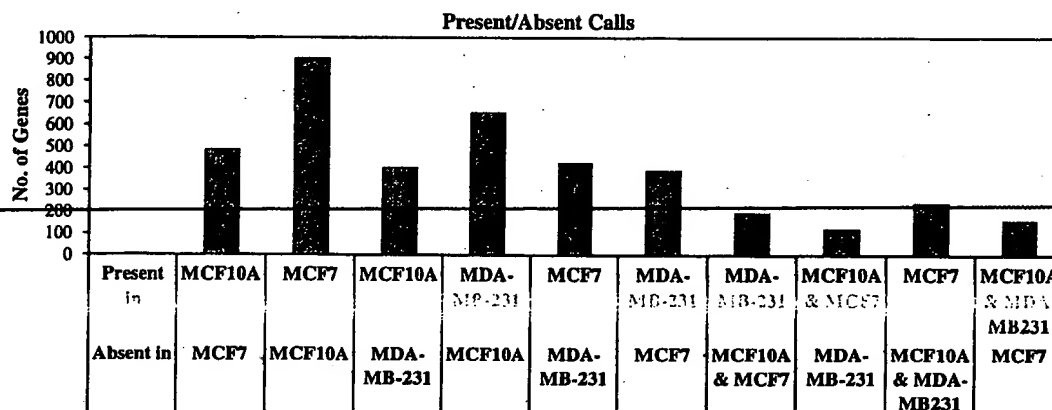


Figure 3 Comparison of cell lines based on the presence or absence of transcripts. The absence or presence of a transcript in the Affymetrix chip was scored by the fluorescence read-out as described in the Materials and methods section.

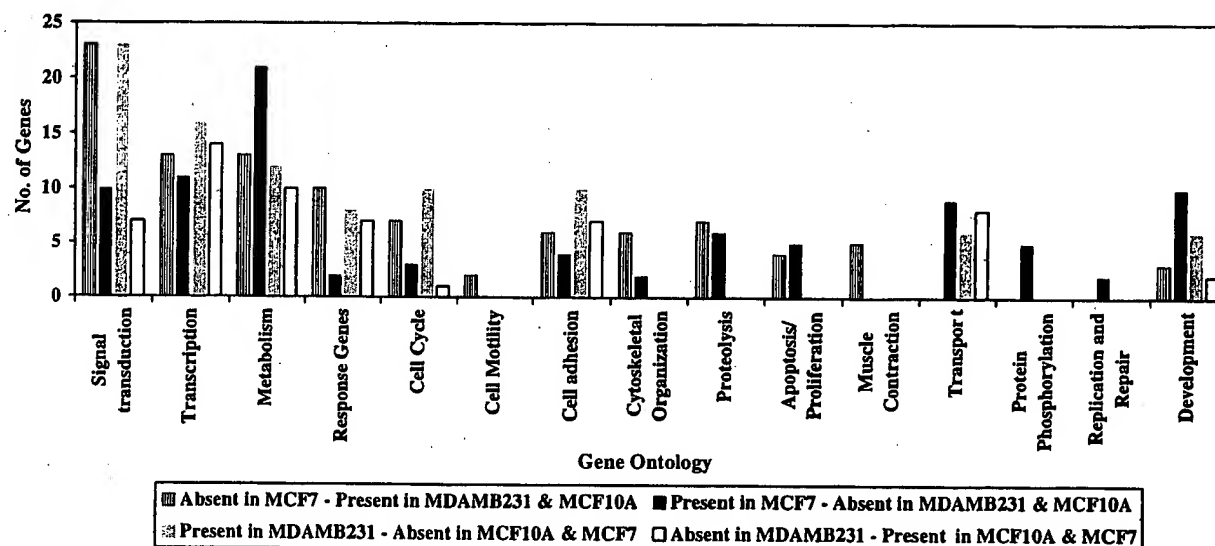


Figure 4 Functional classes of transcripts that differentiate a cell line pair. The transcripts identified as present or absent were classified based on their functional importance.

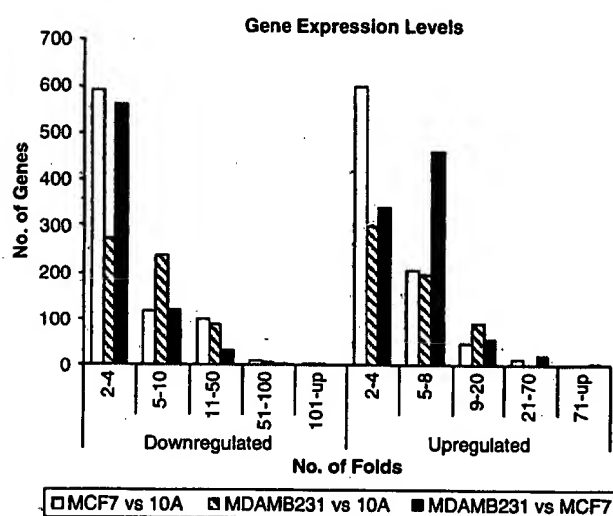


Figure 5 Distribution of altered number of transcripts as a function of fold change. The altered transcripts were categorized in groups based on the magnitude of change in their abundance.

Proteome analysis

To identify altered abundance of proteins and relate it to transcript profiles, we characterized the protein profiles of MCF-10A, MCF-7 and MDA-MB-231 cells. Typically, > 300 protein spots could be visualized in silver-stained gels, and there were far fewer protein spots in gels that were stained with Coomassie blue. The comparison of MCF7 or MDA-MB-231 proteins with MCF-10A revealed that MCF-7 had 11 unique protein spots, while MDA-MB-231 had 15 spots that were not seen in MCF-10A. These proteins were either specific to or upregulated in these cell lines. The identity of these proteins is shown in Table 5. Out of these 26 protein spots, only 25 yielded amino-acid sequence. As shown in the table, the list includes proteins involved in stress

Table 2 Cell adhesion molecules altered in breast carcinoma cells

Upregulated in MDA-MB-231	Downregulated in MDA-MB-231
Cadherin 4	Claudin3
Cadherin 11	Cadherin-1
Catenin	Cadherin-3
Integrin α 6	Cadherin-18
Transmembrane anchor protein	Cadherin, LAG seven pass receptor
Eph B2	Down syndrome cell adhesion
Dystonin	Catenin-32
Laminin β 1	Eph A4
Laminin	Ephrin A4
Filamin B	Annexin A9
Filamin C	Ankyrin 3
Tailin 1	Sarcoglycan
Butyrophilin	Keratin 8
Spectrin- α	MAP-7
Spectrin- β	MAP- τ
Thrombospondin	Plakoglobin
Plastin 3	Plakophilin
Adducin 3 γ	Discoidin domain receptor
Lamin B receptor	Zona occludens 3
Laminin β 2	Periplakin
Lamin A/C	Protocadherin α 9
	Laminin γ 2
	Laminin α 3

response, protein-tagging activities, calcium-binding and calcium homeostasis proteins and some regulatory proteins. Prominent among these changes were proteins involved in calcium homeostasis such as crotactin, calreticulin, calcyclin and reticulocalbin. The changes in signaling pathways between the two cell lines were indicated by altered levels of Rho GDP dissociation inhibitor 1, an apoptosis/differentiation regulating protein galectin, Myc expression regulator far upstream binding protein-1 and the microtubule regulator protein stathmin. The translation initiation factors 5A and 4H were also selectively upregulated in MDA-MB-231 cells.

Table 3 Calcium binding or sensitive transcripts

Upregulated in MDA-MB-231	Downregulated in MDA-MB-231
Reticulocalbin	S100A8
Dystonin	S100A7
Follistatin like 1	S100A13
Cullin	Tumor associated Ca ²⁺ signal transducer
Annexin A5	Notch homolog 3
EF hand domain containing-2	PKD2
Hippocalcin like-2	LDLR
Steroid sulfatase	Adenylate cyclase
MT-actin crosslinking factor	Phospholipase-C
Inositol-1,4,5-triphosphate receptor 3	Chemokine ligand 12
Sorcin	Ubiquitin-specific protease
Guanine nucleotide-binding protein-γ	PKC-η
GAS6	Ret-protooncogene
SWP-70 protein	Signal peptide-CUB domain
Jagged 1	Mannosidase α
EGF-containing fibrulin-like ECM	Bradykinin receptor B2
Transglutaminase 2	Solute carrier family 24
Thrombin receptor-like 1	Ca ²⁺ channel voltage-dependent β3
Plastin 3	Regulator of G protein signaling 17
FYN oncogene	Dystrobin-α
PKC-α	Synaptogamin 1
Calmeglin	Matrix gla protein
Calpain	EF hand domain family member D1
HEG homolog	PK-cyclic AMP dependent
Cyr61	ATPase-Ca ²⁺ transporting
	Calmodulin 1
	CaM kinase

Table 4 Altered transcripts involved in copper homeostasis and Rho signaling

Copper-binding proteins	Rho signaling proteins
LOX	Rho3
LOX-1, LOX-2	Rho/Rac GEF 18
SCO cyt oxidase-deficient homolog 2	Rho GEF 12
COX 17 homolog	Ras related C3 botulinum toxin substrate 2
Metallothionein 1E, 1F and 2a	Cdc 42 effector protein 3
Ring finger protein7	Rho GEF 3
Amiloride-binding protein 1	Rho GDP dissociation inhibitor β
Neurotrypsin/motopsin	RhoD

Changes in proliferation characteristics and protein profiles in response to transfection with antisense constructs of selected transcripts

We had observed significant upregulation of transcripts for DNA replication complex protein GINS PSF2, trefoil factor 3, aurora kinase AIK2, protein kinase H11 and secreted protein XAG in MCF7 cells. We reasoned that antisense knockdown of the above genes in MCF7 cells might indicate pathways involved in tumorigenesis and invasiveness.

MCF7 cells were transfected with empty vector pCDNA3.1 or antisense constructs of the above genes. A semiquantitative amplification of pCDNA marker gene by RT-PCR confirmed the presence of the transfected construct in a significant proportion of the cell population. The transfected cells also showed a decrease in the target transcripts as observed by RT-

Table 5 Proteomic profiles of MCF7 and MDA-MB-231 cells as compared to MCF10A cells

MCF7 cells	MDA-MB-231 cells
<i>Cell-cell adhesion/cell-surface receptor proteins</i>	<i>Calcium-binding proteins</i>
Triose-phosphate isomerase	Calcyclin
	Calreticulin
	Crocalbin
	Reticulocalbin
<i>Stress-response proteins</i>	<i>Transcription Factors/Promoter-binding proteins</i>
Hsp27	Far upstream element binding protein-1
Superoxide dismutase	Far upstream element binding protein-2
Peroxisredoxin 2	
<i>Cytoskeletal component/binding proteins</i>	<i>Cell-cell adhesion/cell-surface receptor proteins</i>
Stathmin	Galectin
<i>Cell-cycle regulation and growth/differentiation/apoptosis proteins</i>	<i>ATPase/GTPase/signal transduction/trafficking proteins</i>
Nucleoside diphosphate kinase A	Rho GDP dissociation inhibitor 1
S100C	
<i>Secreted proteins and growth factors</i>	<i>Stress-response proteins</i>
Macrophage migration inhibitory protein	Hsp 70
	Peroxisredoxin 2
<i>Miscellaneous</i>	<i>Cytoskeletal component/binding proteins</i>
Cyt c oxidase VIb	Stathmin
Peptidyl-prolyl <i>cis-trans</i> isomerase	
Ubiquitin	<i>Miscellaneous</i>
	Heterogeneous nuclear ribonucleo protein H
	eIF4H (translation)
	eIF5A (translation)

PCR. The effects of antisense transfections were scored by growth characteristics of the transfectants. The cell proliferation was reduced between 15 and 40% when antisense transfectants were compared to cells transfected with empty vector.

In order to relate decreased proliferation of transfectants to altered proteins, proteomic profiles of transfectants were compared with vector controls. The comparison of protein profiles of cells transfected with empty vector or antisense construct revealed alterations in several proteins for each transfectant (Table 6). These proteins included stress-response proteins, calcium-regulating proteins, translation factors, ubiquitin, proteins of electron transport chain and oxidative phosphorylation, signaling proteins, cytokeratins, actin and actin regulating proteins and general regulatory factors. The number of altered proteins varied between 5 and 15 for various transfections. Peptidyl prolyl *cis-trans* isomerase, calcium-regulating proteins, SOD, galectin, histidine triad protein and PKC substrate were prominent among altered proteins. We performed database searches to identify interactors for all proteins that were altered in transfected cells and identified nearly 350 proteins (data not shown). A significant number of these interacting proteins are involved in transcriptional regulation.

Table 6 Altered proteins in MCF7 cells after antisense knockdown of specific transcripts

After transfection with as-trefoil factor 3(TFF3)	After transfection with as-protein kinase H11
Calcium-binding proteins Calmodulin Cell-cell adhesion/cell-surface receptor proteins Retinoic acid-binding protein II ATPase-GTPase/signal transduction/trafficking proteins Raf kinase inhibitor Stress-response proteins Hsp27 Peroxiredoxin 1 Cytoskeletal component/binding proteins Cofilin-nonmuscle isoform Cell-cycle regulation and growth/differentiation/apoptosis proteins Chromobox protein homolog 1 Chromobox protein homolog 3 Translationally controlled tumor protein Miscellaneous 40S ribosomal protein S12 ATP synthase D chain Cancer-associated Sm-like protein Cyt c oxidase polypeptide Va eIF5A (translation) Ferritin heavy chain His triad nucleotide-binding protein Histone H2B.n Peptidyl-prolyl isomerase Proteasome subunit β -type 6 RNA-binding protein 8A Thioredoxin Ubiquitin crossreactive protein After transfection with as-aurora kinase AIK2 Calcium-binding proteins Calgranulin B Calgranulin A Stress-response proteins Hsp27 Superoxide dismutase Miscellaneous 60S acidic ribosomal protein P2 Peptidyl-prolyl isomerase Ubiquitin-crossreactive protein	Cell-cell adhesion/cell-surface receptor proteins Galectin Retinoic acid-binding protein II ATPase-GTPase/signal transduction/trafficking proteins PKC substrate Stress-response proteins Hsp27 Cytoskeletal component/binding proteins Actin- α , skeletal muscle Cytokeratin 18 Miscellaneous 3OH-acyl CoA dehydrogenase II ATP synthase A chain Cyt c oxidase peptide Va Enhancer of rudimentary homolog Thioredoxin Ubiquinol-cyt C reductase No match After transfection with as-DNA replication complex GINS PSF2 Calcium Binding proteins: Calreticulin ATPase-GTPase/signal transduction/trafficking proteins 14-3-3 η PKC substrate Cytoskeletal component/binding proteins: Cytokeratin 1 Cytokeratin 18 Cell-cycle regulation and & growth/differentiation/apoptosis proteins Chromobox protein homolog 1 Chromobox protein homolog 3 MAP/MT affinity regulator Miscellaneous ATP synthase D chain Peptidyl-prolyl isomerase Ubiquinol-cyt c reductase After transfection with as-human homolog of XAG Calcium-binding proteins Calreticulin Stress-response proteins Hsp27 Miscellaneous ATP synthase A chain Peptidyl-prolyl isomerase Ubiquitin crossreactive protein No match

Discussion

The results presented here validate the gene profiles obtained from different expression platforms ranging from subtractive hybridization to gene microarrays and proteomic analysis. The RDA protocol is powerful enough to yield important genes that show significant alterations in their expression between cell lines, and can lead to isolation of full-length cDNAs by using appropriate modifications (Baskaran *et al.*, 1996; Jacob *et al.*, 1997). The detection of RUNX2, variant of RUNX2, EphB6, prion protein, lysyl oxidase and a copper transporter ATP7A transcripts by RDA warrant specific mention. RUNX transcription factors bind specific motifs on target gene promoters and regulate gene expression leading to cell growth, proliferation and

differentiation (Pratap *et al.*, 2003). RUNX2 and its variant have differential repression activity toward the promoter of the cyclin-dependent kinase inhibitor (p21CIP1) (Westendorf *et al.*, 2002). The loss of EphB6 expression due to methylation of its promoter is related to invasiveness of MDA-MB-231 (Fox and Kandpal, 2004; unpublished observations). Lysyl oxidase, a copper-sensitive enzyme, causes oxidative deamination of lysine and hydroxy lysines of collagen to aldehyde forms to stabilize collagen fibrils (Siegel, 1976) that are found in invasive breast carcinoma cells (Akiri *et al.*, 2003). The activation of LOX is dependent on copper that is internalized and then transported to trans golgi network by copper transporter ATP7A (Pase *et al.*, 2004), a protein mutated in Menkes disease (Moller *et al.*, 2005). Prion protein has also been reported as a

chaperone for copper (Jones *et al.*, 2004). Thus, EphB6 can serve as a marker for invasiveness, and LOX and ATP7A may be exploited as relevant targets for therapeutic intervention.

The downregulation of junctional proteins along with inactivation of TIMPs as shown here is in agreement with other reports describing their relationship with invasiveness of carcinoma cells (Johnson, 1991; Kousidou *et al.*, 2004; Shao *et al.*, 2005) and promoter methylation (Costa *et al.*, 2004). As several transcripts coding for junctional proteins are downregulated in invasive cells, we postulate that methylation-specific PCR can be exploited to use these transcripts as biomarkers of tumor cells in general and invasiveness in particular. The changes in cell-cell interaction correlate to cell phenotypes because such interactions influence Rho/Ras signal transduction pathways and *vice versa* (Malliri and Collard, 2003; Nagaraja and Kandpal, 2004; Ridley, 2004), and lend credence to the significance of altered transcripts for Rho and Rho GEFs as presented here.

Early changes in calcium homeostasis as measured by calcium excretion have been reported in breast cancer (Campbell *et al.*, 1983), and altered calcium signaling has been shown in invasive lung carcinoma cells (Amuthan *et al.*, 2002). Prominent among calcium-binding proteins are S-100 protein, a group of intracellular messengers that respond to transient changes in calcium concentration by binding to specific receptors (Marenholz *et al.*, 2004) and regulate cell growth, differentiation and motility, transcription and cell cycle. The S-100 proteins detected in the present study map to chromosome 1q21, a region of genome that is frequently altered in human breast cancer cells (Bieche *et al.*, 1995). Calcium ions act as a second messenger in specific signaling pathways in a variety of cancers (Missiaen *et al.*, 2000) and are known to alter calcineurin to activate transcription factors such as NFATc (Luo *et al.*, 1996).

As dictated by post-transcriptional regulation, protein profiles showed far fewer changes as compared to transcript profiles, and the knockdown of five selected genes in MCF7 cells produced interesting changes in protein profiles. These genes, namely, XAG (secretory *Xenopus laevis* protein), trefoil factors 3, human aurora2 kinase AIK2, protein kinase H11 and DNA replication complex GINS PSF2, have been shown to be estrogen responsive, oncogenic or involved in tumorigenesis (Yu *et al.*, 2001; Fletcher *et al.*, 2003; Katoh, 2003; Warner *et al.*, 2003; Takayama *et al.*, 2003). The antisense constructs of these genes appeared to work as siRNAs as suggested by the reduction in the transcript detected in RT-PCR of RNA isolated from the transfected cells. The involvement of the above transcripts in invasive potential is apparent from the observed upregulation of calcium-binding proteins in transfected MCF7 cells, which is comparable to their levels in MDA-MB-231 cells. The proteins that appear to mediate inhibition of proliferation in antisense-transfected cells include PKC substrate, Raf kinase inhibitor, histidine triad nucleotide-binding protein and peptidyl-prolyl isomerase (Pin1). We believe histidine triad protein effects are

most likely mediated via its interaction with ATM protein. Raf kinase inhibitor (Keller *et al.*, 2004) and ATM (Hall, 2005) have been conclusively linked to transformation of cells, and the activity of Pin1 has been related to p53-mediated signaling pathways (Mantovani *et al.*, 2004; Berger *et al.*, 2005). In this context, p53 activation has also been hypothesized by Cu-SOD prion-like enzyme (Wiseman, 2005). Thus, alterations in copper homeostasis and p53-mediated signaling may be considered as a significant regulatory mechanism in tumorigenesis.

In summary, we have presented here a set of candidate genes that can serve as biomarkers for tumorigenesis and invasiveness, and some of these markers may be used to develop DNA-based diagnostic tests. The alterations in transcripts for copper homeostasis genes suggest copper chelation or inhibition of copper transporter ATP7A as potential targets for therapeutic application. The modulation of RUNX2 splicing variants by chemicals that affect splicing machinery may also be explored as a therapeutic modality. The changes in EphB6 expression, if confirmed in tumor specimens, may have prognostic significance.

Materials and methods

Breast cancer cell lines

We used MCF-10A, a cell line established from normal breast, and two breast carcinoma cell lines MCF-7 and MDA-MB-231 that vary in their *in vitro* and *in vivo* invasiveness. All cells were cultured at 37°C/7% CO₂. MCF-10A cells were grown in 1:1 DMEM:F12 media (Gibco) with 5% horse serum (Gibco), 20 mM HEPES, 10 ng/ml EGF (Invitrogen), 10 ml/l PenStrep-Glutamine (10 000 U/ml penicillin, 10 000 µg/ml streptomycin and 29.2 mg/ml L-glutamine), 10 µg/ml insulin (Invitrogen), 0.1 µg/ml Cholera Toxin (Sigma) and 500 ng/ml hydrocortisone (Sigma). MCF-7 and MDA-MB-231 cells were grown in DMEM (Gibco) supplemented with 2 mM L-glutamine (Gibco), 1 mM sodium pyruvate (Gibco), 5 ml/l penstrep (5000 U/ml penicillin and 5000 µg/ml streptomycin), and 10% fetal bovine serum (Hyclone).

Total RNA isolation

RNA was isolated from 85 to 95% confluent 10 cm tissue culture dishes using TRI reagent (Molecular Research Center Inc.) with slight modifications to the recommended protocol. Approximately 10 million cells were mixed with 1.0 ml Tri Reagent, the mixture was extracted with chloroform and the aqueous phase containing RNA was separated. The RNA was precipitated with isopropanol, the pellet washed sequentially with 80 and 100% ethanol, then dried and resuspended in DEPC-treated water. RNA was stored in aliquots at -70°C. The quality of RNA was visualized by running on a formaldehyde gel. The appearance of ribosomal RNA bands indicated that RNA was not degraded during the procedure. The amount of RNA was determined by its absorbance at 260 nm.

DNAase treatment of total RNA

To remove DNA contamination, 20 µg of RNA (quantified spectrophotometrically) was treated with 500 ng DNAase I,

80 U RNasin (Promega) and 1 mM MgCl₂ in Tris buffer in a total volume of 50 μ l. The reaction was carried out at 37°C for 1 h and the DNAase inactivated by heating to 65°C for 30 min.

Representational difference analysis (RDA)

RDA of cDNAs is a modification of genomic RDA (Lisitsyn *et al.*, 1993). We performed RDA in the following two ways. In one experiment, MCF-7 cDNA was used as a driver and MDA-MB-231 cDNA as a tester. In the second experiment, MCF-7 cDNA was used as a tester and MDA-MB-231 cDNA as a driver. The protocol has been described previously (Jacob *et al.*, 1997). Briefly, first-strand synthesis was carried out using a commercial cDNA synthesis kit as per the manufacturer's protocol. A linker with a *Bgl*III site was ligated to the tester as well as the driver cDNA. A primer specific to one of the linker strands was used to PCR amplify the linker-ligated cDNAs. The linkers were then removed by digesting the cDNA with *Bgl*III and the digested cDNA was gel purified. A second set of unphosphorylated *Bgl*III adaptors was ligated to the tester cDNA only. The tester and driver DNAs were hybridized in a 5 μ l reaction volume at a ratio of 1:40. After hybridization, the ends of the tester homoduplexes were repaired with Klenow polymerase and 1 μ l of the reaction mixture was diluted to 100 μ l. The difference product was obtained by amplifying 1 μ l of the diluted mixture using the top strand of the ligated adaptor as a primer. The amplified difference product was digested with *Eco*RI and cloned in a pBlueScript vector. Individual clones were picked up and sequenced by Sanger's dideoxy chain termination method. Representative clones were validated by Northern analysis and semiquantitative RT-PCR.

Microarray analysis

The GeneChips, Human Genome U133A 2.0, (Affymetrix, Santa Clara, CA, USA) used in this study contained approximately 22 000 probe sets corresponding to 18 400 transcripts and variants, including 14 500 well-characterized human genes.

Total RNA was converted into double-stranded cDNA by using SuperScript II (Invitrogen, Carlsbad, CA, USA) and an oligo-dT primer containing a heel of the T7 RNA polymerase promoter sequence. The reaction mixture containing double-stranded cDNA was extracted with phenol-chloroform, precipitated with ethanol, and dissolved in 12 μ l RNase-free water. The cDNA was transcribed *in vitro* by using a RNA transcription labeling kit (Enzo Biochem, Farmingdale, NY, USA) with 6 μ l of double-stranded cDNA in the presence of ATP, CTP, GTP, UTP, bio-11-CTP and bio-16-UTP. The biotinylated RNA was purified by using an affinity column (Qiagen, Valencia, CA, USA) and fragmented randomly, by heating to 95°C in the presence of fragmentation buffer, between sizes of 35 and 200 bases. The GeneChips were hybridized overnight at 45°C in hybridization oven in a solution containing fragmented cRNA, control oligonucleotide B2, 20 \times eucaryotic hybridization controls, herring sperm DNA, acetylated BSA and 2 \times hybridization buffer. The GeneChips were washed and stained with streptavidin-phycoerythrin and the antibody in 2 \times MES stain buffer, acetylated BSA, and optically read at a resolution of 6 μ m with a Affymetrix GeneChip scanner 3000. Affymetrix MICROARRAY SUITE was used for initial data preparation (generation of .CHP files). Normalization (quantile method) and calculation of signal intensities was performed with the software package RMA from the R project (<http://www.r-project.org/>). For every cell line, three replicates were performed with Affymetrix Gene Chips. The Gene Chip data were used for further calculations after the raw image and

MAS5 analysis revealed a positive quality report. Ratios of average signal intensity (log 2) were calculated for the probe sets between pairs of cell lines and then converted to an average fold change (AFC). Statistical validation was performed on probe sets as described (Yoshida *et al.*, 2004). The statistical method used to assign *P*-values to the fold changes of gene responses is described by Yoshida *et al.* (2004) and is a two-step procedure based on the Benjamini and Yekutieli construction of false discovery rate confidence intervals (FDRCI) (Reiner *et al.*, 2003). Functional annotation of proteins was assigned through Gene Ontology (<http://www.geneontology.org>) or Locuslink (<http://www.ncbi.nlm.nih.gov/LocusLink>) classifications obtained through appropriate public databases.

Quantitative RT-PCR

RNA was reverse transcribed with SuperScript II (Invitrogen, Carlsbad, CA, USA) RT by priming with oligodT. The primers specific to validated genes were synthesized from the 3' untranslated region using Primer 3 software. PCR reactions were then performed in triplicates in an I-cycler Thermocycler with optical module (BioRad, Hercules, CA, USA). The amplified products were quantified by reading fluorescence of SybrGreen I (Molecular Probes, Eugene, OR, USA). Average fold changes were calculated by differences in threshold cycles (*C_t*) between pairs of samples to be compared. *HPRT* gene was used as a control.

Semiquantitative RT-PCR

The spectrophotometrically determined concentration of RNA was confirmed by amplifying actin message at different cycles. The cycling conditions that yielded proportional increment of amplified product was used to normalize the RNA concentration. The normalized RNA was used as template to determine relative abundance of transcripts corresponding to clones identified by RDA experiments. The conditions were standardized in the range of cycles that yielded a PCR product for at least one of the pairs of compared RNAs. Such experiments dictated cycles between 30 and 35 to be appropriate to compare abundance of selected transcripts in MCF7 and MDA-MB-231 cells.

Northern analysis

The expression pattern of selected transcripts in cell lines was also analysed in Northern blots. RNA (20 μ m), as determined by spectrophotometer and confirmed by actin amplification, was electrophoresed on a formaldehyde agarose gel. A RNA ladder was used as a size marker. The RNAs were transferred from the gel to a Hybond nylon membrane by capillary transfer. The RNA was fixed onto the membrane by irradiation in a Stratalinker. The blot was hybridized at 65°C for 12–15 h with a radioactive probe and the blot was subsequently washed with 0.1 \times SSC and 1% SDS at 65°C. The hybridized probe was detected by autoradiography.

Transfection of MCF-7 cells with antisense constructs

The genes selected on the basis of their upregulation were cloned in antisense orientation in pcDNA3.1 vector (Invitrogen). MCF-7 cells were grown to 70–80% confluence. Approximately 4 μ g of DNA was transfected into MCF-7 cells by using Lipofectamine 2000. The transfected cells were grown in the presence of G418 (400 μ g/ml). The transfectants were processed for protein isolation. A control set of cells was transfected with an empty pcDNA3.1. The proteins were analysed by two-dimensional electrophoresis, and altered bands were excised for mass spectrometry.

Protein isolation

The cultured cells were harvested by trypsinization and centrifuged at 220 g for 5 min at 4°C. The cell pellet was washed once with ice-cold 1 × PBS. The proteins were isolated by using a commercial kit (BioRad, Hercules, CA, USA). Briefly, pelleted cells (0.05 ml) were mixed with 0.5 ml ice-cold CPEB solution containing protease inhibitors cocktail (Roche), vortexed and stored on ice for 30 min. The cell suspension was passed through a syringe needle (20 gauge) for 10–20 strokes to ensure complete cell lysis. The cytoplasmic protein fraction was collected by centrifugation at 100 g for 10 min at 4°C. The nuclear pellet was washed once again with 0.5 ml CPEB solution. The nuclear pellet was resuspended in 0.75 ml PSB buffer, vortexed briefly and centrifuged at 1000 g for 10 min at 4°C, and the supernatant containing nuclear protein was collected into a new tube. The samples were quantified using 2D Quant kit (Amersham Biosciences), aliquoted and stored at –80°C to prevent protein degradation. To reduce streaking, background staining and the other gel artefacts associated with substances contaminating 2D/IEF samples, the samples were cleaned with 2D Clean up kit (Bio Rad, Hercules, CA, USA) before running on the gel.

Two-dimensional gel electrophoresis

The protein mixtures were separated based on isoelectric points by using commercial pre-cast pH gradient gel strips according to the manufacturer's instructions. The protein sample (175 µg) in 185 µl of sample buffer (8 M urea, 2% CHAPS, 0.2% biolytes, 3/10 ampholytes, 65 mM DTT and 0.002% bromophenol blue) was loaded in the sample loading trays at the end of 11 cm immobilized rehydrated strips (pH 3–10) (Bio Rad, Hercules, CA, USA). Following isoelectric focusing, proteins were reduced and alkylated by successive 15 min treatments with equilibration buffer (6 M urea, 0.375 M Tris-HCl pH 8.8, 2% SDS, 20% glycerol, 2% DTT) and 2.5% (W/V) iodoacetamide, respectively. Proteins were then resolved in the second dimension on 8–16% gradient SDS-PAGE gel (Bio Rad, Hercules, CA, USA). The protein spots were visualized by staining with either silver stain or

Coomassie blue stain. The gel images were compared and bands showing significant (greater than twofold) alterations in intensity were excised and processed for mass spectrometry. Comparisons were made between protein lysates from MCF-10A, MCF-7 and MDA-MB-231 cell lines or between MCF-7 and MCF-7 cells transfected with specific antisense constructs.

Protein identification by enzymatic digestion followed by mass spectrometry

Prior to performing trypsin digestion, the gel pieces containing protein spots were washed sequentially once with water and twice with acetonitrile. The gel pieces were then allowed to swell in 100 mM ammonium bicarbonate and finally washed with acetonitrile. The washed slices were dried in a Speed Vac concentrator, and subsequently incubated with 20 µl of Promega's autocatalysis-resistant trypsin (12.5 ng/µl in 50 mM ammonium bicarbonate and 5 mM CaCl₂ pH 8.0) overnight at 37°C. The supernatant (5 µl) from tryptic digest was injected for peptide sequence analysis using on-line capillary liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-MS/MS). The front end HPLC utilized a Dionex (San Francisco, CA, USA) Vydac 300 µm inner diameter × 15 mm C18 column. The linear acetonitrile gradient (3%/min, containing 0.02% TFA) was developed using a Hewlett-Packard 1100 pump operating at 0.1 ml/min, and the flow was split before the injector such that the flow rate through the column was 3 µl/min. Peptides were detected at 215 nm. The in-line mass spectrometer was a ThermoElectron LCQ-DECA instrument operated in data-dependent MS/MS mode, and proteins were identified by searching a nonredundant protein database using the Sequest program.

Acknowledgements

This research was supported in part by Faculty Research Award, Ames Faculty Award and a grant from the Department of Defense (RPK). We thank Dr M Hamilton for helpful discussions on proteomics.

References

- Akiri G, Sabo E, Dafni H, Vadasz Z, Kartvelishvili Y, Gan N et al. (2003). *Cancer Res* 63: 1657–1666.
- Alpan RS, Sparvero S, Pardee AB. (1996). *Mol Med* 2: 469–478.
- Amuthan G, Biswas G, Ananadatheerthavarada HK, Vijaya-sarathy C, Shephard HM, Avadhani NG. (2002). *Oncogene* 21: 7839–7849.
- Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM. (1996). *Genome Res* 6: 633–638.
- Becker M, Sommer A, Kratzschmar JR, Seidel H, Pohlenz HD, Fichtner I. (2005). *Mol Cancer Ther* 4: 151–168.
- Berger M, Stahl N, Del Sal G, Haupt Y. (2005). *Mol Cell Biol* 25: 5380–5388.
- Bieche I, Champeme MH, Lidereau R. (1995). *Clin Cancer Res* 1: 123–127.
- Campbell FC, Blamey RW, Woolfson AM, Elston CW, Heston DJ. (1993). *Pr J Surg* 71: 202–204.
- Cerosaletti KM, Shapero MH, Fournier RE. (1995). *Genomics* 25: 226–237.
- Costa FF, Verbisck NV, Salim AC, Ierardi DF, Pires LC, Sasahara RM et al. (2004). *Oncogene* 23: 1481–1488.
- Dan S, Tsunoda T, Kitahara O, Yanagawa R, Zembutsu H, Katagiri T et al. (2002). *Cancer Res* 62: 1139–1147.
- Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA et al. (2003). *Br J Cancer* 88: 579–585.
- Fox BP, Kandpal RP. (2004). *Biochem Biophys Res Commun* 318: 882–892.
- Hall J. (2005). *Cancer Lett* 227: 105–114.
- Jacob AN, Baskaran N, Kandpal G, Narayan D, Bhargava AK, Kandpal RP. (1997). *Somatic Cell Mol Genet* 23: 83–95.
- Johnson JP. (1991). *Metast Rev* 10: 11–22, Review.
- Jones CE, Abdelraheim SR, Brown DR, Viles JH. (2004). *J Biol Chem* 279: 32018–32027.
- Katoh M. (2003). *Int J Mol Med* 12: 3–9.
- Keller ET, Fu Z, Yeung K, Brennan M. (2004). *Cancer Lett* 207: 131–137.
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F et al. (2001). *Nat Med* 7: 673–679.
- Koukoudou OC, Roussidis AE, Theodoris AD, Karamanias NK. (2004). *Anticancer Res* 24: 4025–4030.
- Lisitsyn N, Lisitsyn N, Wigler M. (1993). *Science* 259: 946–951.
- Luo C, Shaw KT, Raghavan A, Aramburu J, Garcia-Cozar F, Perrino BA et al. (1996). *Proc Natl Acad Sci USA* 93: 8907–8912.

- Malliri A, Collard JG. (2003). *Curr Opin Cell Biol* 15: 583–589.
- Mantovani F, Gostissa M, Collavin L, Del Sal G. (2004). *Cell Cycle* 3: 905–911.
- Marenholz I, Heizmann CW, Fritz G. (2004). *Biochem Biophys Res Commun* 322: 1111–1122.
- Missiaen L, Robberecht W, van den Bosch L, Callewaert G, Parys JB, Wuytack F et al. (2000). *Cell Calcium* 28: 1–21.
- Moller LB, Bukrinsky JT, Molgaard A, Paulsen M, Lund C, Tumer Z et al. (2005). *Hum Mutat* 26: 84–93.
- Nagaraja GM, Kandpal RP. (2004). *Biochem Biophys Res Commun* 313: 654–665.
- Pase L, Voskoboinik I, Greenough M, Camakaris J. (2004). *Biochem J* 378: 1031–1037.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al. (2000). *Nature* 406: 747–752.
- Pratap J, Galindo M, Zaidi SK, Vradii D, Bhat BM, Robinson JA et al. (2003). *Cancer Res* 63: 5357–5362.
- Reiner A, Yekutieli D, Benjamini Y. (2003). *Bioinformatics* 19: 368–375.
- Ridley AJ. (2004). *Breast Cancer Res Treat* 84: 13–19.
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI et al. (2002). *N Engl J Med* 346: 1937–1947.
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P et al. (2000). *Nat Genet* 24: 227–235.
- Shao Q, Wang H, McLachlan E, Veitch GI, Laird DW. (2005). *Cancer Res* 65: 2705–2711.
- Siegel RC. (1976). *J Biol Chem* 251: 5786–5792.
- Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H. (2003). *Genes Dev* 17: 1153–1165.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al. (2002). *Nature* 415: 530–536.
- Warner SL, Bearss DJ, Han H, Von Hoff DD. (2003). *Mol Cancer Ther* 2: 589–595.
- Westendorf JJ, Zaidi SK, Cascino JE, Kahler R, van Wijnen AJ, Lian JB et al. (2002). *Mol Cell Biol* 22: 7982–7992.
- Wiseman A. (2005). *Med Hypotheses* 65: 32–34.
- Yoshida S, Mears AJ, Friedman JS, Carter T, He S, Oh E et al. (2004). *Hum Mol Genet* 13: 1487–1503.
- Yu YX, Heller A, Liehr T, Smith CC, Aurelian L. (2001). *Int J Oncol* 18: 905–911.

Anuradha Waghray¹
 Farhana Feroze¹
 Megan S. Schober¹
 Fay, Yao¹
 Chris Wood²
 Eric Puravs²
 Melissa Krause²
 Samir Hanash²
 Yong Q. Chen^{1,3}

¹Department of Pathology,
 Wayne State University,
 Detroit, MI, USA

²Department of Pediatrics,
 University of Michigan,
 Ann Arbor, MI, USA

³Center for Molecular Medicine
 and Genetics, Wayne State
 University, Detroit, MI, USA

Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis

A common therapy for nonorgan-confined prostate cancer involves androgen deprivation. To develop a better understanding of the effect of androgen on prostatic cells, we have analyzed gene expression changes induced by dihydrotestosterone (DHT) in the androgen responsive prostate cancer line LNCaP, at both RNA and protein levels. Changes at the RNA level induced by DHT were determined by means of serial analysis of gene expression (SAGE), and protein profiling was done by means of quantitative two-dimensional polyacrylamide gel electrophoresis. Among 123 371 transcripts analyzed, a total of 28844 distinct SAGE tags were identified representing 16570 genes. Some 351 genes were significantly affected by DHT treatment at the RNA level ($p < 0.05$), of which 147 were induced and 204 repressed by androgen. In two independent experiments, the integrated intensity of 32 protein spots increased and 12 decreased at least two-fold in response to androgen, out of a total of 1031 protein spots analyzed. The change in intensity for most of the affected proteins identified could not be predicted based on the level of their corresponding RNA. Our study provides a global assessment of genes regulated by DHT and suggests a need for profiling at both RNA and protein levels for a comprehensive evaluation of patterns of gene expression.

Keywords: Androgen / Prostate / LNCaP / Gene expression / Serial analysis of gene expression
 PRO 0137

1 Introduction

Androgens affect numerous aspects of prostate biology including development, growth, and maintenance. They also play a critical role in tumorigenesis and progression of prostate cancer. Androgen deprivation is an established treatment modality for prostate cancer. However, the disease eventually progresses into a hormone refractory cancer. Several mechanisms have been identified which may contribute to androgen independence: (1) Mutations in the androgen receptor (AR) lead to ligand-independent activation or promiscuity of the receptor [1, 2]. The ability of the receptor to activate or repress downstream genes can also be affected by mutation. AR mutation is associated with advanced phases of prostate cancer [3, 4]; (2) AR can be activated in a ligand-independent manner by specific growth factors and cytokines [5]; (3) AR gene amplification has been found to occur in

28–30% of tumors that recurred post androgen-ablation therapy [6, 7]. Wallen *et al.* [8] have recently shown that one-third of locally recurrent hormone refractory prostate cancer contain AR gene amplification; (4) Coactivator amplification and corepressor down-regulation have been shown to increase receptor transactivation [9, 10]. Regardless of which pathway(s) is taken by the tumor cells, androgen-regulated genes may ultimately be the key players in the development of hormone refractory cancer. As a step towards a better understanding of the effect of androgen on gene expression, we have undertaken a comprehensive assessment of gene expression changes induced by dihydrotestosterone (DHT) in the androgen responsive prostate cancer cell line LNCaP. An important feature of our study is the parallel assessment of expression changes at both RNA and protein levels.

2 Materials and methods

2.1 Prostate cell line and DHT treatment

LNCaP human prostate cancer cell line (American Type Culture Collection, Rockville, MD, USA) was cultured for 3 d in phenol-free RPMI 1640 supplemented with 5% charcoal-stripped fetal bovine serum (FBS) at 37°C. Half of the cultures were then treated with 10^{-9} M DHT for 24 h. Total RNA

Correspondence: Dr. Y. Q. Chen, Department of Pathology, Wayne State University, 540 E. Canfield, Detroit, MI 48201, USA
 E-mail: yqchen@med.wayne.edu
 Fax: +1-313-5770057

Abbreviations: DHT, dihydrotestosterone; G3PD, glyceraldehyde-3-phosphate dehydrogenase; PSA, prostate specific antigen; RT-PCR, reverse transcription-polymerase chain reaction; SAGE, serial analysis of gene expression

and protein were extracted from untreated and DHT treated cells using TRIzol (Invitrogen, Carlsbad, CA, USA) reagent for RNA and a solubilization cocktail for proteins consisting of 9 M urea, β -mercaptoethanol and 1% NP-40. For kinetics experiments 1×10^6 cells were plated in phenol-free RPMI medium with 5% charcoal-stripped FBS. On day three, cells were treated with 5 mL of media containing 10^{-9} M DHT and total RNA was extracted at different time points from 0–96 h. To determine the direct/indirect effect of androgen on RNA levels, cells were plated as above and treated with 5 μ g/mL of cyclohexamide in the presence or absence of 10^{-9} M DHT for 24 h and total RNA was extracted.

2.2 Serial analysis of gene expression

Serial analysis of gene expression (SAGE) was performed as described previously [11] with the following modifications: ditags were PCR amplified using biotinylated primers and digested with *Nla*III enzyme [12]. Concatemers were heated for 15 min at 65°C and chilled on ice for 10 min before separation on an 8% polyacrylamide gel [13]. The concatemers were then cloned into the SphI site of the pZero vector (Invitrogen). Concatenated tags were screened by PCR using M13 forward and M13 reverse primers. PCR products with inserts greater than 500 bp were isolated and sequenced with M13 forward primer on an automated 3700 DNA sequencer (Perkin-Elmer, Norwalk, CT, USA). For microSAGE, 1 μ g of total RNA per tube was used for cDNA synthesis in two tubes, with the mRNA Capture Kit (Boehringer Mannheim, Indianapolis, IN, USA). cDNAs were cleaved with *Nla*III, ligated to linkers and then digested with *Bsm*FI enzyme. The released tags were ligated, and processed for the rest of steps as with the standard SAGE protocol.

2.3 SAGE data analysis

SAGE tags were extracted using the SAGE software V 4.12 [11]. Tags were matched to the SAGE reliable map (release 10–26–2000) (<http://www.ncbi.nlm.nih.gov/SAGE/>). Due to the fact that some tags map to multiple genes and some genes have multiple tags, SAGE data were analyzed in two different ways: (1) exclusion method: tags that match to multiple genes were discarded. Only tags that match to a single gene were tabulated and composite counts analyzed for their significance; (2) inclusion method: tags that match to multiple genes were counted at 100% toward each gene. All tags were tabulated and composite counts analyzed for their statistical significance. Lists of differentially expressed genes ($p < 0.05$) obtained from the exclusion and inclusion methods were compared, and finally only genes that have a p value < 0.05 in both lists were considered statis-

tically significant. The total number of genes identified was estimated by $N_m + (N_{um} - 0.1 N_{um})/3.5$, where N_m is the number of genes matched to SAGE tags, N_{um} is the number of SAGE tags that do not match to known genes or ESTs, with 10% representing the estimated sequencing error per SAGE tag and 3.5, the average number of tags per gene in the SAGE reliable map (release 10–26–2000).

2.4 RT-PCR and real-time PCR quantification

For reverse transcription-PCR (RT-PCR), 1 μ g of total RNA was reverse transcribed into cDNA. One fortieth of cDNA was used for PCR reaction. Samples were collected at different cycles and separated on a 2% agarose gel with ethidium bromide. Image was captured and quantified using NucleoVision 760 Image Workstation (Nucleotech, CA, USA). Amplification curves were generated. Subsequently, RT-PCR was done at cycles within the log phase of amplification.

Real-time quantification was performed in the iCycler (Bio-Rad, Hercules, CA, USA). Briefly, one fortieth of cDNA was used in each reaction. Six reactions were carried out for each gene and three independent experiments were performed. PCR mix comprised of 1X PCR buffer, 1.5 mM $MgCl_2$, 0.1 mM dNTP, 200 nM primers (listed below), 0.05 U platinum Taq polymerase (Invitrogen) and 0.1x SYBR green (Molecular Probes, Eugene, OR, USA). PCR was carried out at 94°C for 2 min, and 40 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Primers used for real-time PCR quantification: PSA (prostate specific antigen) (Hs.171995, 5'-GGAAATGACCAGGCCAAGAC-3', 5'-CAACCCTG GAC C TCACACCTA-3'), SCMH1 (Hs.57475, 5'-GCCTTGACC ACATCACTCCAT-3', 5'-AGGCCTAGGGCTGCAAAAG-3'), and clusterin (Hs.75106, 5'-GCAGGAATACCGCAAAAA GC-3', 5'-GACTCAAGATGCCCCCGTAAG-3'). Standard samples (50, 25, 10, 5, 2.5, 1 and 0.5 μ L cDNA) and experimental samples were used in real-time quantification PCR. Each sample was run in quadruple reactions. Standard curve ($C_t = mX + B$) was obtained, where C_t is threshold cycle number, X is log quantity of target molecules, m is curve slope and B is Y-axis intercept value. Number of fold induction or repression for a given target molecule was calculated by Q_a/Q_b , where Q_a is the target quantity in experimental sample A and Q_b is the target quantity in experimental sample B. $C_{ta} = m \log Q_a + B$ or $Q_a = 10^{(C_{ta}-B)/m}$, therefore, $Q_a/Q_b = 10^{(C_{ta}-C_{tb})/m}$.

2.5 2-D PAGE

The procedure followed was as previously described [14]. Cells were solubilized in 200 μ L of lysis buffer containing 9.5 M urea (Bio-Rad), 2% NP-40, 2% carrier

ampholytes pH 4–8 (Gallard/Schlessinger, Carle Place, NY, USA), 2% β -mercaptoethanol and 10 mM PMSF. Aliquots containing approximately 5×10^5 cells, were applied onto isofocusing gels. IEF was conducted using pH 4 to 8 carrier ampholytes at 700 V for 16 h, followed by 1000 V for an additional 2 h. The 1-D gel was loaded onto the 2-D gel, after equilibration in 125 mM Tris, pH 6.8, 10% glycerol, 2% SDS, 1% DTT and bromophenol blue. For the second dimension separation, a gradient of 11–14% of acrylamide (Serva, Hauppauge, NY, USA) was used. Following 2-D PAGE, proteins were visualized by silver staining of the gels or transferred to an Immobilon-P PVDF membrane (Millipore, Bedford, MA).

2.6 Protein identification by mass spectrometry

The 2-D gels were silver stained by successive incubations in 0.02% sodium thiosulfate for 2 min, 0.1% silver nitrate for 40 min and 0.014% formaldehyde plus 2% sodium carbonate. The proteins of interest were excised from the 2-D gels and destained for 5 min in 15 mM potassium ferricyanide and 50 mM sodium thiosulfate as described [15]. Following three washes with water, the gel pieces were dehydrated in 100% acetonitrile for 5 min and dried for 30 min in a vacuum centrifuge. Digestion was performed by addition of 100 ng of trypsin (Promega, Madison, WI, USA) in 200 mM ammonium bicarbonate or by addition of 100 ng of the endoproteinase Glu-C (Promega) in 100 mM ammonium bicarbonate. The Lys-C digestion was performed with 500 ng of the endoproteinase Lys-C (Roche, Mannheim, Germany) in 100 mM Tris-HCl, pH 9. Following enzymatic digestion overnight at 37°C, the peptides were extracted twice with 50 μ L of 60% acetonitrile/1% TFA. After removal of acetonitrile by centrifugation in a vacuum centrifuge, the peptides were concentrated by using pipette tips C18 (Millipore).

Analyses were performed primarily using a PerSeptive Biosystem MALDI-TOF Voyager-DE mass spectrometer (Framingham, MA, USA), operated in delayed extraction mode. Peptide mixtures were analyzed using a saturated solution of cyano-4-hydroxycinnamic acid (CHA) (Sigma, St. Louis, MO, USA) in acetone containing 1% TPA (Sigma). Peptides were selected in the mass range of 800–4000 Da. Spectra were calibrated using calibration mixture 2 of the Sequazyme peptide mass standards kit (PerSeptive Biosystems). The search program MS-Fit, developed by the University of California at San Francisco (<http://prospector.ucsf.edu>), was used for searches in the NCBI database. Search parameters were as follows: maximum allowed peptide mass error of 400 ppm, consideration of one incomplete cleavage per peptide and

pH range between 4 and 8. MALDI-TOF mass spectrometry was also used for molecular weight determination as described [16]. In some cases, the amino acid sequence of some peptides of interest was determined by ESI MS analysis.

3 Results

3.1 SAGE analysis of the effect of androgen on gene expression

SAGE libraries were generated from LNCaP cells cultured in the presence or absence of DHT. A total of 123 371 tags were generated of which 62 878 were from the LNCaP cell line and 60 493 from the LNCaP cells treated with DHT for 24 h (Table 1). Sequence analysis identified a total of 28 844 distinct tags representing 16 570 genes, 11 243 from LNCaP and 12 203 from DHT treated cells. A total of 351 transcripts were differentially expressed at a significant level ($p < 0.05$). Eighty-seven percent of transcripts matched GenBank entries; 79% corresponded to known sequences and 8% to ESTs. RNA levels for 147 genes were increased and 204 genes were decreased after DHT treatment (Table 2). Therefore, at the RNA level more genes were repressed than activated by androgens. Of these androgen-regulated genes, 149 were changed \geq five-fold by DHT treatment (Table 3).

Table 1. Summary of SAGE analysis in LNCaP cells

Sample	Total transcripts	Distinct tags	Number of genes
– DHT	62 878	17 050	11 243
+ DHT	60 493	18 510	12 203
Total	123 371	28 844	16 570

– DHT: LNCaP cells without dihydrotestosterone; + DHT: LNCaP cells treated with 10^{-9} M dihydrotestosterone for 24 h

Table 2. DHT regulated genes in LNCaP cells

Genes	LNCaP vs LNCaP + DHT (gene #)	
	$p < 0.05$	$p < 0.01$
Known genes	277	147
ESTs	29	8
No match	45	13
Total	351	168
Up-regulated by DHT (24 h)	147	65
Down-regulated by DHT (24 h)	204	103

Table 3. List of candidate genes that are regulated by DHT (≥ 5 fold)

UGD or Tag	CAP	DHT	Change	P value	Description
283305	1	82	Induced	0	immunoglobulin heavy constant alpha 1
183752	0	42	Induced	0	microseminoprotein, beta-
140	1	44	Induced	0	immunoglobulin heavy constant gamma 3 (Gm marker)
75415	18	95	Induced	0	beta-2-microglobulin
84298	0	17	Induced	1.00E-05	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)
9615	0	16	Induced	1.00E-05	myosin regulatory light chain 2, smooth muscle isoform
77443	0	16	Induced	1.00E-05	actin, gamma 2, smooth muscle, enteric
1119	0	15	Induced	2.33E-05	nuclear receptor subfamily 4, group A, member 1
84753	23	2	Repressed	3.00E-05	KIAA0246 protein
75777	0	14	Induced	3.66E-05	transgelin
	0	14	Induced	3.66E-05	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides
76325					
1852	1	17	Induced	4.33E-05	acid phosphatase, prostate
78465	0	13	Induced	9.00E-05	v-jun avian sarcoma virus 17 oncogene homolog
78344	1	16	Induced	1.20E-04	myosin, heavy polypeptide 11, smooth muscle
83006	19	2	Repressed	1.50E-04	CGI-139 protein
263812	15	1	Repressed	3.19E-04	nuclear distribution gene C (A. nidulans) homolog
TACGGGGATA	0	11	Induced	4.20E-04	Novel
GCCTGGGTGG	11	0	Repressed	6.63E-04	Novel
GACTGACACT	16	2	Repressed	8.40E-04	Novel
284296	0	10	Induced	8.60E-04	Homo sapiens SURF-4 mRNA, complete cds
75105	13	1	Repressed	1.19E-03	emopamil-binding protein (sterol isomerase)
128075	10	0	Repressed	1.30E-03	ESTs
154162	15	2	Repressed	1.43E-03	ADP-ribosylation factor-like 2
6895	1	12	Induced	1.49E-03	actin related protein 2/3 complex, subunit 3 (21 kD)
143240	17	3	Repressed	1.63E-03	ESTs
211582	0	9	Induced	1.66E-03	myosin, light polypeptide kinase
180266	0	9	Induced	1.66E-03	tropomyosin 2 (beta)
93002	14	2	Repressed	2.53E-03	ubiquitin carrier protein E2-C
17883	14	2	Repressed	2.53E-03	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform
126023	14	2	Repressed	2.53E-03	ESTs, Highly similar to NTC1_HUMAN NEUROGENIC-LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR
69469	2	13	Induced	2.92E-03	dendritic cell protein
77899	2	13	Induced	2.92E-03	tropomyosin 1 (alpha)
285501	2	13	Induced	3.07E-03	Human rearranged immunoglobulin lambda light chain mRNA
119209	0	8	Induced	3.07E-03	insulin-like growth factor binding protein 7
173043	0	8	Induced	3.07E-03	metastasis-associated 1-like 1
75866	0	8	Induced	3.07E-03	dimethylarginine dimethylaminohydrolase 1
171695	0	8	Induced	3.07E-03	dual specificity phosphatase 1
TACGGGGATT	0	8	Induced	3.34E-03	Novel
172791	13	2	Repressed	4.19E-03	ubiquitously-expressed transcript
13561	13	2	Repressed	4.19E-03	ESTs, Weakly similar to dJ37E16.5 [H. sapiens]
98260	8	0	Repressed	4.34E-03	ESTs
CATAAGACTT	8	0	Repressed	4.64E-03	Novel
TACGGGGACA	2	12	Induced	5.76E-03	Novel
180034	12	2	Repressed	6.91E-03	cleavage stimulation factor, 3' pre-RNA, subunit 3, 77kD
5753	12	2	Repressed	6.91E-03	inosito(myo)-1(4)-monophosphatase 2
278941	10	1	Repressed	6.94E-03	PRO0628 protein
19500	0	7	Induced	7.25E-03	nuclear localization signal deleted in velocardiofacial syndrome

Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
9006	0	7	Induced	7.25E-03	VAMP (vesicle-associated membrane protein)-associated protein A (33kD)
26471	0	7	Induced	7.25E-03	<i>Homo sapiens</i> clone HQ0692
AACTGCTGGC	12	2	Repressed	7.63E-03	Novel
GCTGACCGTC	7	0	Repressed	8.59E-03	Novel
CCCCCTGTC	7	0	Repressed	8.59E-03	Novel
243901	2	11	Induced	8.82E-03	<i>Homo sapiens</i> cDNA FLJ20738 fis, clone HEP08257
119000	1	9	Induced	9.12E-03	actinin, alpha 1
78596	1	9	Induced	9.12E-03	proteasome (prosome, macropain) subunit, beta type, 5
171955	7	0	Repressed	9.61E-03	trophinin associated protein (tastin)
180545	7	0	Repressed	9.61E-03	<i>Homo sapiens</i> mRNA for hypothetical protein (TR2/D15 gene)
77719	7	0	Repressed	9.61E-03	gamma-glutamyl carboxylase
4877	7	0	Repressed	9.61E-03	CGI-51 protein
22795	7	0	Repressed	9.61E-03	ESTs
79335	7	0	Repressed	9.61E-03	<i>Homo sapiens</i> SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin D1 (SMARCD1) mRNA
119177	9	1	Repressed	1.26E-02	ADP-ribosylating factor 3
CGGGAGCACC	9	1	Repressed	1.29E-02	Novel
182217	0	6	Induced	1.32E-02	succinate-CoA ligase, ADP-forming, beta subunit
75106	0	6	Induced	1.32E-02	clusterin (complement lysis inhibitor, SP-40, 40, sulfated glycoprotein 2, testosterone-repressed prostate message 2)
256311	0	6	Induced	1.32E-02	granin-like neuroendocrine peptide precursor
103180	0	6	Induced	1.32E-02	DC2 protein
TACGGGGATG	1	8	Induced	1.70E-02	Novel
AAACAAATCA	2	10	Induced	1.70E-02	Novel
136644	2	10	Induced	1.71E-02	CS box-containing WD protein
8036	2	10	Induced	1.71E-02	glioblastoma overexpressed
74284	2	10	Induced	1.71E-02	ESTs, Weakly similar to Similar to <i>S. cerevisiae</i> hypothetical protein L3111
ATGGCTGATC	6	0	Repressed	1.75E-02	Novel
ATCACTGCCC	6	0	Repressed	1.75E-02	Novel
ACATCATCAG	6	0	Repressed	1.75E-02	Novel
CCAGTCCAAG	6	0	Repressed	1.75E-02	Novel
54842	1	8	Induced	1.78E-02	ESTs
1526	1	8	Induced	1.78E-02	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
7911	1	8	Induced	1.78E-02	KIAA0323 protein
227400	1	8	Induced	1.78E-02	mitogen-activated protein kinase kinase kinase kinase 3
77269	1	8	Induced	1.78E-02	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
7943	6	0	Repressed	1.81E-02	RPB5-mediating protein
278544	6	0	Repressed	1.81E-02	acetyl-Coenzyme A acetyltransferase 2 (acetoacetyl Coenzyme A thiolase)
23111	6	0	Repressed	1.81E-02	phenylalanine-tRNA synthetase-like
31442	6	0	Repressed	1.81E-02	RecQ protein-like 4
38628	6	0	Repressed	1.81E-02	hypothetical protein
77422	6	0	Repressed	1.81E-02	proteolipid protein 2 (colonic epithelium-enriched)
211973	6	0	Repressed	1.81E-02	homolog of Yeast RRP4 (ribosomal RNA processing 4), 3'-5'-exoribonuclease
171075	6	0	Repressed	1.81E-02	replication factor C (activator 1) 5 (36.5kD)
89781	6	0	Repressed	1.81E-02	upstream binding transcription factor, RNA polymerase I
154149	6	0	Repressed	1.81E-02	apurinic/aprimidinic endonuclease(APEX nuclease)-like 2 protein

Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
279772	8	1	Repressed	2.28E-02	brain specific protein
205091	8	1	Repressed	2.28E-02	ESTs, Weakly similar to WW domain binding protein 11 (<i>M. musculus</i>)
75658	8	1	Repressed	2.28E-02	phosphorylase, glycogen; brain
GGGCAGCTGT	8	1	Repressed	2.37E-02	Novel
105440	0	5	Induced	2.43E-02	hepatocyte nuclear factor 3, alpha
118244	0	5	Induced	2.43E-02	protein phosphatase 2, regulatory subunit B (B56), delta isoform
82389	0	5	Induced	2.43E-02	CGI-118 protein
31638	0	5	Induced	2.43E-02	restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)
12013	0	5	Induced	2.43E-02	ATP-binding cassette, sub-family E (OABP), member 1
12797	0	5	Induced	2.43E-02	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
7736	0	5	Induced	2.43E-02	hypothetical protein
274479	0	5	Induced	2.43E-02	NME7
153138	0	5	Induced	2.43E-02	origin recognition complex, subunit 5 (yeast homolog)-like
16034	0	5	Induced	2.43E-02	ESTs
270072	0	5	Induced	2.43E-02	ESTs
187035	0	5	Induced	2.43E-02	ESTs
140452	0	5	Induced	2.43E-02	cargo selection protein (mannose 6 phosphate receptor binding protein)
182265	0	5	Induced	2.43E-02	keratin 19
19762	0	5	Induced	2.43E-02	ESTs, Weakly similar to unknown (<i>D. melanogaster</i>)
317	0	5	Induced	2.43E-02	topoisomerase (DNA)
6236	0	5	Induced	2.43E-02	ESTs
GCTGGAGCCT	5	0	Repressed	3.03E-02	Novel
CCAGTGCTCA	5	0	Repressed	3.03E-02	Novel
ACCCTACATA	5	0	Repressed	3.03E-02	Novel
GGGGAAATCT	5	0	Repressed	3.03E-02	Novel
ACTGGTACTG	5	0	Repressed	3.03E-02	Novel
GCTCCGGTGT	5	0	Repressed	3.03E-02	Novel
ACAGTGGTGA	5	0	Repressed	3.03E-02	Novel
7869	1	7	Induced	3.04E-02	lysophosphatidic acid acyltransferase-delta
12101	1	7	Induced	3.04E-02	hypothetical protein
266940	1	7	Induced	3.04E-02	t-complex-associated-testis-expressed 1-like 1
6196	1	7	Induced	3.04E-02	integrin-linked kinase
366	1	7	Induced	3.04E-02	6-pyruvoyltetrahydropterin synthase
173611	1	7	Induced	3.04E-02	NADH dehydrogenase (ubiquinone) Fe-S protein 2 (49kD) (NADH-coenzyme Q reductase)
102469	1	7	Induced	3.04E-02	putative nuclear protein
78825	1	7	Induced	3.04E-02	matrin 3
284465	1	7	Induced	3.04E-02	ESTs
30738	1	7	Induced	3.04E-02	hypothetical protein FLJ10407
78687	1	7	Induced	3.04E-02	neutral sphingomyelinase (N-SMase) activation associated factor
92381	1	7	Induced	3.04E-02	nudix (nucleoside diphosphate linked moiety X)-type motif 4
242039	5	0	Repressed	3.26E-02	EST
4766	5	0	Repressed	3.26E-02	DKFZP586O0120 protein
283109	5	0	Repressed	3.26E-02	hypothetical protein DKFZp762L1710
192853	5	0	Repressed	3.26E-02	ubiquitin-conjugating enzyme E2G 2 (homologous to yeast UBC7)
153678	5	0	Repressed	3.26E-02	reproduction 8

Table 3. Continued

UGD or Tag	CAP	DHT	Change	P value	Description
251317	5	0	Repressed	3.26E-02	EST
279623	5	0	Repressed	3.26E-02	selenoprotein X
150319	5	0	Repressed	3.26E-02	ESTs
102456	5	0	Repressed	3.26E-02	survival of motor protein interacting protein 1
284250	5	0	Repressed	3.26E-02	AD-003 protein
251871	5	0	Repressed	3.26E-02	CTP synthase
270480	7	1	Repressed	3.92E-02	ESTs, Weakly similar to ALU5_HUMAN ALU SUBFAMILY SC SEQUENCE
26655	7	1	Repressed	3.92E-02	glucose-6-phosphatase, transport (glucose-6-phosphate) protein 1
8118	7	1	Repressed	3.92E-02	leukotriene A4 hydrolase
CTCCGCCGGC	7	1	Repressed	4.47E-02	Novel
AGGAAATGCT	7	1	Repressed	4.47E-02	Novel
GCTGACCGAGG	7	1	Repressed	4.47E-02	Novel
CCCATAGTCC	7	1	Repressed	4.47E-02	Novel

CAP: LNCaP cells without dihydrotestosterone

DHT: LNCaP cells treated with 10^{-8} dihydrotestosterone for 24 h

UGD: unigene ID

Numbers in CAP and DHT columns are the number of tags observed in respective samples

3.2 Confirmation of SAGE data

To confirm the differential expression pattern at the RNA level, semiquantitative RT-PCR assay was performed on a group of selected genes. One μ g of total RNA was reverse transcribed into cDNA with oligo dT. cDNA was amplified by PCR for various cycles. PCR products were separated on agarose gels and quantified by densitometry. PCR amplification curves were plotted for each gene, and data within the logarithmic phase of amplification were used for quantification (Fig. 1A, B and C). The reproducibility of gene regulation by androgen was confirmed in three independent experiments by monitoring the induction of PSA in the presence or absence of cyclohexamide (Fig. 1B). The kinetics of gene induction or repression was determined for PSA (Hs.171995), clusterin (Hs.75106) and SCMH1 (Hs.57475) genes. Typical examples of kinetics are shown in Figure 1C. PSA was induced at 4-6 h, peaked between 6-20 h, and gradually declined after 20 h post-treatment of DHT. Clusterin was induced within 0.5-1 h and gradually declined after 6-12 h. SCMH1 was repressed 2-4 h post-treatment. Expression of PSA, clusterin, and SCMH1 was quantified by real-time PCR (Fig. 1D).

3.3 Changes in protein expression induced by DHT

Protein lysates of LNCaP cells cultured in the presence of DHT for 72 h and cells cultured in parallel in the absence of DHT were subjected to 2-D PAGE (Fig. 2). Following sil-

ver staining, gels were digitized prior to computer-based matching and quantitative analysis, as previously described [17]. Of a total of 1031 protein spots matched between 2-D patterns of DHT treated and untreated cells, a set of 32 protein spots increased in intensity by at least two-fold in two independent experiments. Likewise, a set of 12 protein spots decreased in intensity by at least two-fold in these two independent experiments. A two-fold change in integrated intensity by silver staining represents on average a three-fold change in amount of protein, based on prior quantitative studies [17].

The 44 protein spots that changed in intensity as a result of DHT treatment were excised from the gels, digested with trypsin, and subsequently analyzed by MALDI-TOF MS. The resulting spectra were used to identify the proteins using the MS-Fit search program. Of the 44 spots excised from the gels, 29 were identified without ambiguity and consisted of 21 up-regulated and eight down-regulated proteins (Table 4). The identified proteins represented a heterogeneous group that included cytoskeletal proteins (e.g. tropomyosin, α tubulin), metabolic enzymes (e.g. adenine phosphoribosyl transferase, β 1,4 galactosyl transferase, galactocerebrosidase), and the products of previously described androgen responsive genes (e.g. SRY, selenium binding protein) [18, 19]. Specific antibodies confirmed the identification based on MS for all proteins analyzed by Western blotting, which included α tubulin, myosin light chain isoforms, nucleoside diphosphate kinase A, glyceraldehyde 3-phosphate dehydrogenase (G3PD), and tropomyosin (data not shown).

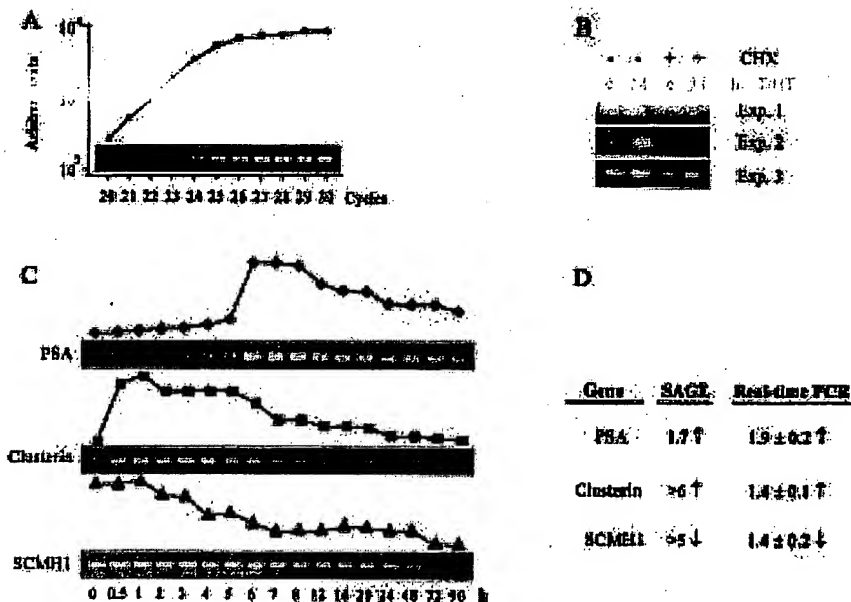


Figure 1. Kinetics and quantification of androgen-regulated genes. Total RNA from untreated and DHT treated LNCaP cells was reverse transcribed. cDNA was PCR amplified for various cycles. PCR amplification curves were plotted for each gene. A cycle number within the logarithmic phase of amplification was selected for semi-quantitative analysis. cDNA was also used for real-time quantitative PCR. (A) An example of RT-PCR amplification of PSA at cycles between 20–30. (B) RT-PCR of PSA in three independently treated LNCaP samples. h: hours; 0: no DHT; 24: treated with 10^{-8} M DHT for 24 h; -: without cycloheximide; +: with 5 μ g/mL of cycloheximide for 24 h. (C) An example of three independent kinetics experiments for PSA, clusterin and SCMH1. (D) Comparison of SAGE results with that of real-time quantitative PCR for PSA, clusterin and SCMH1. ↑: induction; ↓: repression.

without cycloheximide; +: with 5 μ g/mL of cycloheximide for 24 h. (C) An example of three independent kinetics experiments for PSA, clusterin and SCMH1. (D) Comparison of SAGE results with that of real-time quantitative PCR for PSA, clusterin and SCMH1. ↑: induction; ↓: repression.

Corresponding SAGE data were available for most of the proteins affected by DHT treatment that were identified. It was therefore of interest to determine if the changes observed at the protein levels were matched by concordant changes at the mRNA level. Remarkably, for most of the proteins identified, there was no appreciable concordant change at the RNA level (Table 4).

4 Discussion

We have compared the gene expression profile of LNCaP, an androgen responsive prostate cancer cell line, with or without androgen (DHT) treatment. Approximately 350 of 16570 expressed genes detected at the RNA level were affected by dihydrotestosterone at the $p < 0.05$ level. The DHT responsive genes included known genes, ESTs, and novel genes. As expected, we saw an induction of genes that are well known to be regulated by androgens. For instance, we found a 1.7-fold induction in kallikrein 3/PSA, a 7.6-fold induction in prostatic kallikrein 2, and a 15.7-fold induction in prostatic acid phosphatase by DHT in LNCaP cells. We also saw a five-fold induction in NKX3.1/NKX3A and three-fold in fatty acid synthase; two previously identified androgen-regulated genes [20, 21]. Interestingly β -microseminoprotein, reported to be reduced or lost in prostate tumor [22], was up-regulated more than 40-fold in

LNCaP cells by DHT. More significantly, our data indicate that genes involved in a variety of tumor cell functions such as growth, apoptosis, and metastasis, are directly or indirectly regulated by androgens. In addition, it is noteworthy that both β -actin and G3PD, the two most frequently used loading controls, were up-regulated approximately two-fold by DHT ($p < 0.01$). Thus, these two genes may not be appropriate controls in some experiments.

Serial analysis of androgen-regulated gene expression provides us with a list of candidate genes. However, many factors such as the dose of DHT and the time of cell exposure to DHT and other unknown experimental variations will affect the level of gene expression. Therefore, it is important to confirm the SAGE results with alternative methods. We performed semiquantitative RT-PCR (Fig. 1B) on approximately 20 genes and real-time quantitative PCR on 10 genes in at least three independently DHT treated LNCaP samples. We noticed experimental variations in every gene determined. Some genes such as PSA (Fig. 1B) have smaller variations and some have larger variations. In addition, the fold of alteration in expression identified by SAGE is different for some genes compared to that identified by real-time quantitative PCR. This is probably due to the technical limitation of SAGE. Serial analysis of gene expression is highly quantitative for genes whose tags are detected at large numbers.

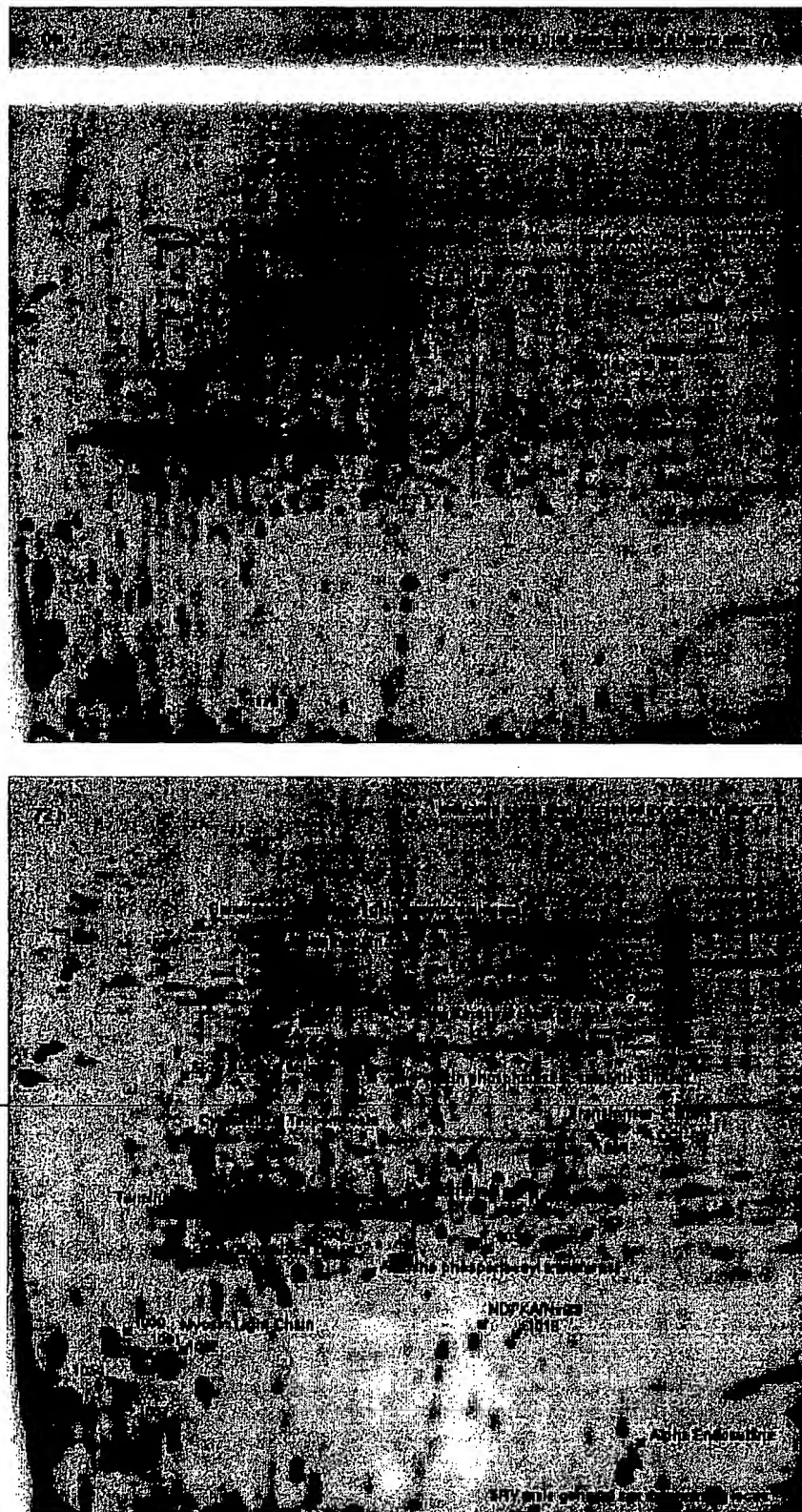


Figure 2. 2-D PAGE patterns of untreated (A) and LNCaP cells treated with DHT for 72 h (B). Identified proteins that were reduced in levels following DHT treatments are shown in (A) and induced proteins are shown in (B). Numbered spots represent either proteins that were changed in level that have not been identified or isoforms of identified proteins.

Table 4. List of proteins that are altered by DHT (≥ 2 -fold)

Protein name or spot number	Unigene ID	Protein		Change in protein	mRNA		Change in mRNA
		0 h	72 h		0 h	24 h	
Adenine phosphoribosyl transferase	28914	0.448	1.084	↑	36	18	↓
Alpha Endosulfine	111680	0.066	0.679	↑	2	5	↑
Alpha Tubulin	278242	0.924	1.932	↑	48	38	↓
Nucleophosmin, B23	173205	0.256	0.522	↑	16	5	↓
CGI-83	118554	0.063	0.291	↑	2	3	↑
Creatine Kinase B chain (brain)	173724	0.508	1.142	↑	59	32	↓
Cytokeratin 8	242463	0.134	0.393	↑	35	40	↑
Cytoskeletal tropomyosin, NTRK1	85844	0.414	1.174	↑	14	18	↑
Galactocerebrosidase	273	0.095	0.405	↑	3	3	=
Lactoyl Glutathione Lyase	75207	0	1.256	↑	27	22	↓
Myosin Light Chain (1000)	77385	0.724	6.102	↑	118	88	↓
Myosin Light Chain (1001)	77385	0.6	3.212	↑	118	88	↓
Myosin Light Chain (1032)	77385	2.573	5.694	↑	118	88	↓
Nm23	118638	0.583	1.978	↑	35	25	↓
Protein phosphatase 6, catalytic Protein	279563	0.287	0.737	↑	2	2	=
	7016	0	0.448	↑	2	5	↑
Slow Skeletal Muscle Troponin	84673	0.067	0.195	↑	1	0	↓
SRX male gonadal sex determining	1992	0	0.735	↑	ND	ND	
Synovial sarcoma, X breakpoint 2	289105	0.225	0.593	↑	ND	ND	
Terminal deoxynucleotidyl	272537	0.392	0.995	↑	ND	ND	
Transformer-2 alpha	119523	0.074	0.201	↑	7	13	↑
284		0.173	0.461	↑			
402		0	0.33	↑			
456		0.154	0.314	↑			
664		0.297	0.629	↑			
807		0	0.377	↑			
863		0.246	0.665	↑			
903		0.271	0.632	↑			
1018		0	0.725	↑			
1054		0.47	1.394	↑			
1057		0.451	0.959	↑			
1082		0.609	1.738	↑			
14-3-3 eta	75544	0.232	0.114	↓	1	5	↑
Beta-1,4-galactosyl transferase	158540	1.306	0.707	↓	ND	ND	
Cox11	241515	1.643	1.453	↓	3	0	↓
Dihydrolipoamide dehydrogenase	74635	0.844	0.132	↓	9	8	↓
G3PD (456)	169476	1.191	0.279	↓	75	137	↑
G3PD (457)	169479	2.051	0.336	↓	75	137	↑
Selenium Binding Protein (234)	7833	0.411	0.201	↓	26	31	↑
Selenium Binding Protein (237)	7833	0.574	0.22	↓	26	31	↑
651		0.31	0.115	↓			
745		0.21	0.067	↓			
892		0.674	0.454	↓			
1178		0.933	0.305	↓			

Numbers in protein are the intensity of protein spots

Numbers in mRNA columns are the number of SAGE tags detected

ND: not detected

↑: increase; ↓: decrease; =: equal

When the tag number detected for a given gene is close to zero, the quantitative nature of SAGE is compromised. For instance, in approximately 60 000 transcripts of each sample, the PSA tag was detected 42 times in LNCaP and 70 times in LNCaP treated with DHT. However, the tag for clusterin was detected 0 times and 6 times, and the tag for SCMH1 5 times and 0 times in the corresponding samples. Therefore, the SAGE data of PSA is more reliable than that of clusterin and SCMH1. Indeed, SAGE data of PSA is virtually identical to that of real-time quantitative PCR, whereas the SAGE data of clusterin and SCMH1 differ from that of quantitative PCR (Fig. 1D).

PSA is probably the best-known androgen-regulated gene. We were surprised to see only 1.7-fold induction by DHT. However, kinetics experiments indicate that PSA is induced at 4–6 h, peaked between 6–20 h, and gradually declined 20 h post-treatment of DHT. Since the SAGE experiment was done in samples treated for 24 h with DHT, PSA mRNA level was likely to have declined from its peak level. Indeed, we could detect 5–10 fold induction of PSA at 6–8 h post-treatment. Clusterin was reported as an androgen-repressed gene in the rat prostate [23]. Recent evidence indicates that clusterin may not be directly androgen-repressed, but regulated by apoptotic stimuli [24]. Our results suggest that clusterin was induced within 0.5–1 h, gradually declined after 6–12 h, and after 24 h reduced to a lower level than that of untreated cells (Fig. 1C).

Another important question is how many of the androgen-regulated genes identified are directly induced or repressed by androgens. In order to address this question, detailed analyses must be done for each gene. As the first step, we will divide the androgen-regulated genes identified in this study into two groups; the cyclohexamide-sensitive group whose induction or reduction of expression by DHT is blocked by the protein synthesis inhibitor and the cyclohexamide-insensitive group whose induction or reduction is not affected by cyclohexamide.

We consider that genes in the latter group are directly regulated by DHT. Preliminary results suggest that approximately 20% of the 20 genes studied by RT-PCR are cyclohexamide-insensitive. Experiments are under way to determine the cyclohexamide sensitivity of all of the 149 genes listed in Table 3. Further characterization of androgen-regulated genes may provide some clues on the transition from hormone sensitive to hormone refractory prostate cancer.

A relatively small set of genes could be analyzed at the protein level, largely due to the limited sensitivity of 2-D PAGE. Nevertheless, a substantial number of detected proteins (44 of 1031 proteins analyzed, 4.3%) were affected by DHT treatment. Some of the proteins in this

subset that were identified included the products of genes that were previously shown to be affected by androgens, namely selenium binding protein, brain specific creatine kinase, SRY protein, B23 and G3PD. Using a subtractive approach, the human selenium-binding protein gene was shown to be differentially expressed in LNCaP and reversibly down-regulated by exogenous androgen in a concentration-dependent manner, in concordance with our findings at the protein level [19]. An increase in levels of the brain specific creatine kinase B chain in response to androgen has also been described [25]. Likewise the SRY gene has been found to be responsive to androgen stimulation in LNCaP cells [18]. Also, a number of glycolytic enzymes including G3PD which were affected at the protein level have been found to be responsive to androgen stimulation [26]. Androgenic regulation of the amount and phosphorylation of the protein B23, included in our list of affected proteins, has been previously described and found to be related to the early changes associated with androgen mediated growth of the prostate [27].

Corresponding SAGE data was available for most of the proteins affected by DHT treatment that were identified. Interestingly, for most of the proteins identified, there was no appreciable concordant change at the RNA level. There are several potential explanations for this lack of concordance. A particular gene may be represented by more than one protein isoform in 2-D gels. For example, in the identified group, three proteins (myosin light chain, G3PD and selenium binding protein) were represented by more than one isoform. Thus, a source of discordance between RNA and protein data may be that the protein change is limited to a particular isoform of a protein and not to overall protein products of a particular gene. Nevertheless, a change in a particular isoform is informative and biologically meaningful and may not be predictable from RNA data. A lack of concordance between RNA and protein data may also reflect either translational control, post-translational modifications, or changes in protein turnover due to DHT treatment. Yet another explanation for a lack of concordance could be a lag time for changes at the RNA level to be reflected in a protein change. It follows from the above considerations that monitoring gene expression at both RNA and protein levels may provide complementary information that could not be ascertained by solely measuring RNA or protein.

We would like to thank Dr. Ken Kinzler (Johns Hopkins University) for providing SAGE software. This study is partially supported by DAMD17-98-1-8501, R01CA74927 and a fund from Karmanos Cancer Institute.

Received May 21, 2001

5 References

- [1] Taplin, M. E., Bubley, G. J., Shuster, T. D., Frantz, M. E., et al., *N. Engl. J. Med.* 1995, 332, 1355–1361.
- [2] Elo, J. P., Kvist, L., Leinonen, K., Isomaa, V., et al., *J. Clin. Endocrinol. Metab.* 1995, 80, 3494–3500.
- [3] Taplin, M. E., Bubley, G. J., Ko, Y. J., Small, E. J., et al., *Cancer Res.* 1999, 59, 2511–2515.
- [4] Marcelli, M., Ittmann, M., Mariani, S., Sutherland, R., et al., *Cancer Res.* 2000, 60, 944–949.
- [5] Weigel, N. L., Zhang, Y., *J. Mol. Med.* 1998, 76, 469–479.
- [6] Visakorpi, T., Hyytinen, E., Koivisto, P., Tanner, M., et al., *Nat. Genet.* 1995, 9, 401–406.
- [7] Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., et al., *Cancer Res.* 1997, 57, 314–319.
- [8] Wallen, M. J., Linja, M., Kaartinen, K., Schleutker, J., Visakorpi, T., *J. Pathol.* 1999, 189, 559–563.
- [9] Bautista, S., Valles, H., Walker, R. L., Anzick, S., et al., *Clin. Cancer Res.* 1998, 4, 2925–2929.
- [10] Smith, C. L., Nawaz, Z., O'Malley, B. W., *Mol. Endocrinol.* 1997, 6, 657–666.
- [11] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, 270, 484–487.
- [12] Powell, J., *Nucleic Acid Res.* 1998, 26, 3445–3446.
- [13] Kenzelmann, M., Muhlemann, K., *Nucleic Acid Res.* 1999, 27, 917–918.
- [14] Strahler, J. R., Julick, R., Hanash, S. M., in: Creighton, T. E. [Ed.] *Protein Structure: A Practical Approach*, IRL Press, England 1989, pp. 65–92.
- [15] Ghosh, S., R. V. A. Chapp, C. R., Monaghan, D. A., Inzel, P. S., Mische, S. M., *Electrophoresis* 1999, 20, 601–605.
- [16] Liang, X., Bai, J., Liu, Y.-H., Lubman, D. M., *Anal. Chem.* 1996, 68, 1012–1018.
- [17] Hanash, S. M., in: Hames, B. D., [Ed.], *Gel Electrophoresis of Proteins*, IRL Press, Oxford, 1998, pp. 189–211.
- [18] Lau, Y. F., Zhang, J., *Mol. Carcinog.* 2000, 27, 308–321.
- [19] Yang, M., Sytkowsk, A. J., *Cancer Res.* 1998, 58, 3150–3153.
- [20] Prescott, J. L., Blok, L., Tindall, D. J., *Prostate* 1998, 35, 71–80.
- [21] Swinnen, J. V., Ullrich, W., Heyns, W., Verhoeven, G., *Proc. Natl. Acad. Sci. USA* 1997, 94, 12975–12980.
- [22] Chan, P. S., Chan, L. W., Xuan, J. W., Chai, H. L., Chan, F. L., *Prostate* 1999, 41, 99–109.
- [23] Léger, J. G., Montpetit, M. L., Tenniswood, M., *Biochem. Biophys. Res. Commun.* 1987, 147, 196–203.
- [24] Miyake, H., Nelson, C., Rennie, P. S., Gleave, M. E., *Cancer Res.* 2000, 60, 170–176.
- [25] Somjen, D., Lundgren, S., Kaye, A. M., *J. Steroid Biochem. Mol. Biol.* 1997, 62, 89–96.
- [26] Tawfic, S., Goueli, S. A., Olson, M. O., Ahmed, K., *Cell. Mol. Biol. Res.* 1993, 39, 43–51.
- [27] Brooks, D. E., *Biochem. J.* 1976, 156, 527–537.

REGULAR ARTICLE

Web-based data warehouse on gene expression in human colorectal cancer

Emil Sagynaliev¹, Ralf Steinert², Gerd Nestler², Hans Lippert², Manfred Knoch¹ and Marc-André Reymond²

¹ Department of Surgery, Johanniter Krankenhaus, Stendal, Germany

² Department of Surgery, University of Magdeburg, Magdeburg, Germany

Based on biomedical literature databases, we tried a first step for constructing a gene expression "data warehouse" specific to human colorectal cancer (CRC). Results of genome-wide transcriptomic research were available from 12 studies, using various technologies, namely, SAGE, cDNA and oligonucleotide arrays, and adaptor-tagged amplification. Three studies analyzed CRC cell lines and nine studies of human samples. The total number of patients was 144. Out of 982 up- or down-regulated genes, 863 (88%) were found to be differentially expressed in a single study, 88 in two studies, 22 in three studies, 7 in four studies, and only 2 genes in six studies. Eight large-scale proteomics studies were published in CRC, using 2-D-, SDS- or free-flow electrophoresis, involving only 11 patients. Out of 408 differentially expressed proteins, 339 (83%) were found to be differentially expressed only in a single study, 16 in three studies, 10 in four studies, 3 in five, and 1 in eight studies. Confirmation at proteome level of results obtained with large-scale transcriptomics studies was possible in 25%. This proportion was higher (67%) for reproducing proteome results using transcriptomics technologies. Obviously, reproducibility and overlapping between published gene expression results at proteome and transcriptome level are low in human CRC. Thus, the development of standardized processes for collecting samples, storing, retrieving, and querying gene expression data obtained with different technologies is of central importance in translational research.

Received: December 27, 2004

Revised: March 26, 2005

Accepted: March 29, 2005

Keywords:

Colon neoplasms / Rectal neoplasms / Transcriptomics

1 Introduction

Translational research in human colorectal cancer (CRC) is applying a large spectrum of molecular biology, cellular biology, and advanced validation tools. In particular, genome-wide techniques are now applied to decipher modifications

in gene expression. In recent years, transcriptomics and proteomics tools have been broadly applied in CRC. The hope is that the new data obtained will now allow a classification of disease on molecular basis, deep insights into the pathophysiology of CRC, prognostic statements, and finally a systematic search for diagnostic and therapeutic targets.

Because of the complexity of the biological system under investigation, the most significant contribution of translational research in CRC is expected to derive not from the analysis of single experiments but from libraries of experiments. In other words, the results obtained so far by translational research tools in different clinical and experimental settings need to be compared, contrasted, and if possible synthesized. Thus, based on the biomedical literature databases, we tried a first step for constructing such "data warehouse" specific to human CRC.

Correspondence: Professor Marc-André Reymond, University of Magdeburg, Chairman and Head, Department of Surgery, Evangelic Hospital Bielefeld, Burgsteig 13, D-33617 Bielefeld, Germany

E-mail: marc.reymond@ukb.uni-magdeburg.de
Fax: +49-521-144-4551

Abbreviations: cDNA, complementary DNA; CRC, colorectal cancer

The specific aims of the present study were to screen transcriptomics and proteomics studies published in CRC, and to elaborate a kind of "meta-analysis of gene expression" in CRC. Our hypothesis was that we would be able to determine a common set of genes and gene products that are up- or down-regulated in human CRC by overlapping results obtained by different authors.

2 Materials and methods

This is a literature survey of gene expression data published in human CRC. A Medline search of reports published in English from 1990 to 2004 using the terms "colon cancer" and "gene expression" with the limit "human" was performed and yielded 1979 articles. All abstracts were reviewed and a related article search was performed on appropriate abstracts. Articles were selected by a consensus of two reviewers (E.S., M.R.) that satisfied these predetermined criteria: sample origin (human and/or cell lines) and preparation detailed, technology for gene expression studies defined (transcriptomics and/or proteomics studies), and quantitative results on over-expression or underexpression available. Studies concerning single genes or arbitrarily selected genes were discarded. Results obtained in animal models were not considered.

Data were entered into an Excel working sheet (Microsoft, Seattle, USA). Since gene expression data were not always obtained with quantitative, but in most cases semi-quantitative gene expression analysis technologies, no expression values or ratios were entered into the database. Only genes reported as being over- or underexpressed were entered. No threshold was defined so that some genes defined as differentially expressed might have shown only

marginal differences. Only descriptive statistics are provided, which were obtained with the in-built tools of the Excel software.

Various methods were used for large-scale translational research in CRC, including transcriptomics and proteomics technologies. Results of genome-wide transcriptomic research are available from 12 studies, using various technologies, namely, SAGE (five studies), complementary DNA (cDNA) arrays (five studies), and oligonucleotide array (one study), as well as adaptor-tagged amplification (one study). Three studies analyzed CRC cell lines and nine other studies analyzed human samples. The total number of patients profiled in all studies was 144. Out of the nine studies with human samples, only two studies with a total of only 22 patients were performed using purified epithelial cells (one of them using cDNA arrays, the other one using SAGE technology). An overview of these studies is provided in Table 1.

Eight large-scale proteomics studies were published in CRC using various technologies, namely, 2-D PAGE (six studies), SDS-PAGE (one study), and free-flow electrophoresis (one study). In total, all proteome studies in CRC involved only 11 patients. An overview of these studies is provided in Table 2.

3 Results

One thousand two-hundred and forty genes have been reported to be dysregulated (up- and/or down-regulated) in human CRC, representing about 5% of the 20 000–25 000 human genes.

* The complete dataset can be consulted:
<http://www.chirurgiebethel.de>

Table 1. Large-scale or genome-wide transcriptomics studies in human CRC, using cell lines or human tissues. All studies together involved only 144 patients

Author	Year	Sample	Preparation	Method	Number of patients
Zhang <i>et al.</i> [18]	1997	Normal mucosa and primary tumors	Whole tissue	SAGE	2
SAGE-NET [19]	1997	Primary tumors and cell lines	Cell lines	SAGE	–
Zhang <i>et al.</i> [18]	1997	Cell lines	Cell lines	SAGE	–
Parle-McDermott <i>et al.</i> [20]	2000	Cell lines	Cell lines	SAGE	–
Yanagawa <i>et al.</i> [21]	2001	Paired normal and cancer	Whole tissue	cDNA	20
Takemasa <i>et al.</i> [22]	2001	Paired normal and primary tumors	Whole tissue	cDNA	20
Nottermann <i>et al.</i> [23]	2001	Paired normal mucosa and primary tumors, or adenoma	Whole tissue	Oligonucleotide array	22
Buckhaults <i>et al.</i> [24]	2001	Normal mucosa, adenoma and primary tumors	Purified epithelial cells	SAGE	6
Birkenkamp <i>et al.</i> [25]	2002	Paired normal and primary tumors	Whole tissue	cDNA	27
Lin <i>et al.</i> [26]	2002	Paired normal mucosa and primary tumors, or adenoma	Purified epithelial cells	cDNA	16
Muro <i>et al.</i> [27]	2003	Normal and cancer	Whole tissue	Adaptor-tagged	11
Williams <i>et al.</i> [28]	2003	Paired normal mucosa and primary tumors, normal and adenoma	Whole tissue	cDNA	20

Table 2. Large-scale or proteomics studies in human CRC, using cell lines or human tissues. All studies together involved only nine patients

Author	Publication year	Sample	Method	Number of patients
Reymond <i>et al.</i> [9]	1997	Purified epithelial cells	2-D PAGE	1
Simpson <i>et al.</i> [29]	2000	Cell lines	SDS-PAGE	–
Lawrie <i>et al.</i> [30]	2001	Purified epithelial cells	2-D PAGE	4
Simpson <i>et al.</i> [31]	2001	Cell lines	Free-flow electrophoresis	–
Medjahed <i>et al.</i> [32]	2003	<i>In silico</i>	2-D PAGE	–
Demalte <i>et al.</i> [33]	2003	Cell lines	2-D PAGE	–
Stierum <i>et al.</i> [34]	2003	Cell lines	2-D PAGE	–
Friedman <i>et al.</i> [35]	2004	Human whole tissue	2-D PAGE	6

The vast majority of dysregulated genes in human CRC was found using transcriptomics tools: a total of 982 genes was found to be differentially expressed in at least one of these 12 transcriptomics studies. Out of these 982 genes, 863 (88%) were found to be differentially expressed only in a single study, in other words these results have not been reproduced so far. The other findings could be reproduced in two or more transcriptomics studies: 88 genes were found to be differentially expressed in two studies, 22 in three studies, 7 in four studies, no gene in five studies, and 2 genes in six studies. The most cited genes are listed in Tables 3 (up-regulation) and 4 (down-regulation).

A total of 408 proteins were found to be differentially expressed in human CRC in at least one study. Out of these 408 molecules, 339 (83%) were mentioned in a single study, in other words these results could not be reproduced. Differential expression of the remaining 70 proteins were mentioned with following frequency: 40 were mentioned in two proteomics studies, 16 in three studies, 10 in four studies, 3 in five studies, and a single protein in eight studies. The most cited proteins are listed in Table 5. It has to be noted that only a single study [1] provided differential display protein expression data obtained in the human patient, using whole tissue biopsy.

It is also difficult to reproduce transcriptomics results with proteomics tools. Out of 982 genes found to be differentially expressed in human CRC by genome-wide transcriptomics technologies (Table 6a), only 177 (18%) have been confirmed using proteome technologies. When the genes reported to be differentially expressed in three and more transcriptomics studies ($n = 31$) are compared with those reported to be differentially expressed in three and more proteomics studies ($n = 30$), only two genes (actin: ACTB_HUMAN and creatin-kinase: KCRB_HUMAN) can be matched. Thus, the probability of reproducing a gene expression result obtained at transcriptome level is low when using proteomics technologies.

In contrast, there is a better reproducibility when proteomics results are verified using transcriptomics tools (Table 6b). In fact, when the subset of 30 proteins that are consistently (in three studies or more) reported to be dysregulated in CRC is compared with transcriptomics results, 20

genes can be matched, representing about two-thirds. Thus, the probability of being able to reproduce a proteomics result using transcriptomics tools is more than three times higher than the opposite way (67 vs. 18%). However, it has to be noted that in 16% of cases, results remain unclear or even contradictory.

Both in the transcriptomics and proteomics studies, many genes and factors were found to be differentially regulated that obviously do not play a causal role in CRC carcinogenesis.

4 Discussion

This study was aimed at screening transcriptomics and proteomics studies published in CRC, and to elaborate a kind of meta-analysis of gene expression in CRC in order to generate a data warehouse that would be useful in translational research. A significant number of translational research studies have been published in CRC, starting in 1997 with SAGE technology, followed by an increasing number of array-based transcriptomics studies since 2001, and more recently by several proteomics studies. This has created a significant amount of data which we were then able to compile.

Our starting hypothesis was that we would be able to determine a common set of genes and gene products that are up- or down-regulated in human CRC by comparing results obtained by different authors. This endeavor was not very successful; obviously, overlapping between published gene expression results both at proteome and transcriptome level is low in human CRC. In fact, more than 80% of results could not be reproduced. Several explanations can be provided to explain this lack of reproducibility.

First, the number of patients included in the studies is low (altogether 144 patients in transcriptomics, 11 in proteomics studies); some high impact publications having been conducted on samples of only two patients [18]. This is a problem because interindividual genetic variability is high in human CRC [2]. Thus, it is allowed to hypothesize that some results attributed to gene dysregulation might be caused by genetic diversity rather than by cancer-specific traits.

Table 3. Results of transcriptomics research in human CRC. Most cited up-regulated genes out of 12 genome-wide or large-scale studies

				TRANSCRIPTOMICS											
Accession Number	Swiss-Prot	Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
AF041260	O75363	IFM2_HUMAN	Interferon-inducible protein 1-Bd	1		1	1			1					
AA232508	Q9Y397	SAHH_HUMAN	S-Adenosylhomocysteine hydrolase AHcy			1	1				1	1			
AA007218	Q96QY8	BGH3_HUMAN	Transforming growth factor, beta-induced, 66 kDa	1		1	1				1				
G01119		TCPD_HUMAN	Chaperonin containing TCP-1, subunit 4 (delta)					1				1	1		
X75821	Q04984	CA11_HUMAN	COL1A1				1	1		1					
U22055	Q96AG0	TF3A_HUMAN	General transcription factor IIIA				1	1						1	
U14631	P80365	GRO_HUMAN	GRO1 oncogene (melanoma growth stimulating activity, alpha)				1					1	1		
M92439	P42704	MDP1_HUMAN	MPDP4, MDP7 microsomal dipeptidase			1	1			1					
AF029082	P31947	MPI2_HUMAN	M-phase inducer phosphatase 2	1	1	1									
M86400	P29312	PROC_HUMAN	Pyrroline 5-carboxylate reductase mRNA	1	1	1									
L76465	P15428	SPRC_HUMAN	Secreted protein, acidic, cysteine-rich (osteonectin)	1	1	1									
X57352	Q01628	Q96AG0	100 kDa coactivator mRNA	1	1										
X03205	P62888	RL30_HUMAN	50S Ribosomal protein L30	1	1										
M95787	Q01995	RL23_HUMAN	60s ribosomal protein L23	1	1										
Z25821	P42126	HS9B_HUMAN	90-kDa HSP	1	1										
U96132	Q99714	RLA2_HUMAN	Acidic ribosomal phosphoprotein P2 mRNA	1	1										
AB028893	P04643	CA14_HUMAN	Alpha-1 chain of collagen IV	1	1										
X79239	Q02546	APA1_HUMAN	Apolipoprotein A-I	1	1										
M13934	P06366	BMP4_HUMAN	Bone morphogenetic protein-2B (BMP-2B)	1	1										
M60854	P17008	CSE1_HUMAN	CAS chromosome segregation gene homologue	1	1										
M13932	P06708	CD81_HUMAN	CD81 antigen	1	1										
X69150	P25232	TCPZ_HUMAN	Chaperonin containing TCP1, subunit 6A(zeta1)	1	1										
L06498	P17075	CA21_HUMAN	Collagen alpha-2 type 1	1	1										

Table 3. Continued

				TRANSCRIPTOMICS											
Accession Number	Swiss-Prot	Entry Name	Name												
				cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
L04483	P35265	CO2_HUMAN	Complement component C2	1	1										
M64716	P25111	R10A_HUMAN	Csa-19mRNA	1	1										
X77770	P02383	ROA1_HUMAN	DNA binding protein UPI, liver mRNA fragment	1	1										
W52460	P14798	TP2A_HUMAN	DNA topoisomerase II (top2)	1	1										
X55715	P23396	IF39_HUMAN	Eukaryotic translation initiation factor(EIF3)mRNA	1	1										
M77234	P49241	IF2B_HUMAN	Eukaryotic translation initiation factor 2, subunit 2	1	1										
M58458	P12750	GGH_HUMAN	Gamma-glytamyl hydrolase	1	1										
D16992	P10660	G3P2_HUMAN	Glyceraldehyde 3 phosphate dehydrogenase	1	1										
M77233	P23821	SYG_HUMAN	Glycyl-tRNA synthetase	1	1										
F16294	P09058	RS8_HUMAN	H19 RNA	1	1										
X61156	P08865	HS9A_HUMAN	Heat shock Protein HSP 90-Alpha	1	1										
L36055	Q13541	HMG1_HUMAN	Hmg1 mRNA for high mobility group protein I	1	1										
AA316619	P04645	RS23_HUMAN	Human homolog of yeast ribosomal protein S23 (D14530)	1	1										
L14599	Q15233	IMD2_HUMAN	IMP (inosine monophosphate) dehydrogenase 2	1	1										
X79234	P39026	IFM1_HUMAN	Interferon induced transmembrane protein 1(9-27)	1	1										
L06505	P30050	LDHB_HUMAN	Lactate dehydrogenase B(LDH-B)	1	1										
X64707	P26373	RS2_HUMAN	LLRep 3	1	1										
X56932	P40429	4F2_HUMAN	Lymphocyte activation antigen 4F2 large subunit	1	1										
L25899	P39030	NPL1_HUMAN	Nucleosome assembly protein 1-like 1	1	1										
X63527	P14118	GSHH_HUMAN	Phospholipid hydroperoxide glutathione peroxidase	1	1										
X89401	P46778	GDFH_HUMAN	Prostate differentiation factor	1	1										
H59771	P23131	RL6_HUMAN	Ribosomal protein L 6	1	1										
M94314	P38663	RL18_HUMAN	Ribosomal protein L18(RPL18) mRNA	1	1										
L19527	P08526	RL1X_HUMAN	Ribosomal protein L18a mRNA, complete cds	1	1										
U14968	P46776	RL7_HUMAN	Ribosomal protein L7	1	1										
U14969	P46779	RL8_HUMAN	Ribosomal protein L8	1	1										
L38941	P49207	RS5_HUMAN	Ribosomal protein S5	1	1										
U12465	P42766	RS19_HUMAN	S 19 ribosomal protein	1	1										
F19234	P18077	MYC_HUMAN	V-myc myelocytomatosis viral oncogene homolog (avian)	1	1										
X66699	P12751	Q99497	EST(PARK7; DJ1)	1	1										

Table 4. Results of transcriptomics research in human CRC. Most cited down-regulated genes out of 12 genome-wide or large-scale studies

		TRANSCRIPTOMICS											
Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array
CAH2_HUMAN	Carbonic anhydrase II	1		1	1	1				1		1	1
CEA1_HUMAN	EST (BILIARY GLYCOPROTEIN)			1	1	1				1		1	1
BENE_HUMAN	BENE	1			1				1				1
CAH1_HUMAN	Carbonic anhydrase I	1		1						1			1
GUAA_HUMAN	GCAP-II guanylate cyclase activator 2B; Uroguanylin; UGN	1			1					1			1
FAB1_HUMAN	L-FABP liver fatty acid-binding protein 1	1			1	1				1			
ACDS_HUMAN	Acyl-CoenzymeA dehydrogenase C-2 to C-3 short chain				1							1	1
ADHG_HUMAN	ADH gamma 2 subunit (aa-1–375)				1	1							1
T4S3_HUMAN	Co-029; transmembrane 4 superfamily member 3			1	1					1			
DRA_HUMAN	Colon mucosa-associated (DRA)				1					1			1
KCRB_HUMAN	Creatine kinase	1			1					1			
GUAN_HUMAN	GUCA1B guanylate cyclase activator 1B/guanylin				1					1			1
O95784	IgG Fc-binding protein	1			1	1							
ITMC_HUMAN	Integral membrane protein 2C					1	1			1			
MK03_HUMAN	MAPK3				1					1		1	
FXY3_HUMAN	MAT8 protein				1	1				1			
ACTB_HUMAN	mRNA fragment encoding cytoplasmic actin	1						1	1				
SELP_HUMAN	Selenoprotein P				1	1					1		
EST1_HUMAN	Liver carboxylesterase 1 [Precursor]									1	1		1
CFAD_HUMAN	Adipsin, complement factor DCDA (EST)				1								1
ATPB_HUMAN	ATP5B ATP synthase				1	1							
CAHC_HUMAN	CA 12				1	1							
CALN_HUMAN	Calmodulin-dependent protease (small subunit)				1					1			
CALM_HUMAN	Calmodulin-1 (CALM1) mRNA, 3' UTR, partial sequence								1			1	
CAH4_HUMAN	Carbonic anhydrase IV				1								1

Table 4. Continued

		TRANSCRIPTOMICS												
Entry Name	Name	cDNA	Adaptor-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array	
MT1E_HUMAN	cDNA similar to gb: M10942_cds1 human								1				1	
O00748	CES 2 carboxylesterase 2				1	1								
CMGA_HUMAN	CgA				1								1	
CLUS_HUMAN	clusterin												1	
K1CT_HUMAN	Cytokeratin 20								1				1	
KDGA_HUMAN	Diacylglycerol kinase				1								1	
DTD_HUMAN	DTD sulfate transporter				1	1								
PLA8_HUMAN	EST 122594 5_								1				1	
ATPA_HUMAN	F1-ATPase alpha subunit				1	1								
LEG3_HUMAN	Galectin-3 (Galactose-specific lectin 3) (MAC-2 antigen)				1	1								
ABP_HUMAN	HP-DAO1 (diamine oxidase)										1		1	
K1CS_HUMAN	Keratin 19						1				1			
K2C8_HUMAN	Keratin, type II cytoskeletal 8 (cytokeratin 8) (K 8) (CK 8)							1			1			
DHB2_HUMAN	L11708 Estradiol-17 beta- dehydrogenase 2						1						1	
MT1H_HUMAN	Metallothionein 1H												1	
MUC2_HUMAN	Mucin 2, intestinal/tracheal							1	1					
MEPA_HUMAN	PPH alpha gene											1	1	
PA2A_HUMAN	RASF-A PLA2 gene						1						1	
MT1F_HUMAN	RNA helicase-related protein												1	
SBP1_HUMAN	SBP selenium-binding protein						1	1						
MT1L_HUMAN	Serine theonine kinase 39 (STE20/SPS1 homolog,yeast)						1						1	
MYL6_HUMAN	Myosin light polypeptide 6									1		1		
TETN_HUMAN	TNA tetranectin						1						1	
VIPR_HUMAN	VIPR1 vasoactive intestinal polypeptide receptor 1						1						1	
PLA8_HUMAN	Placenta-specific gene 8 protein											1	1	
Q9NXM9	Hypothetical protein FLJ20151						1						1	
T4S1_HUMAN	Transmembrane 4 superfamily member 1				1							1		
ALU2_HUMAN	Alu subfamily SB sequence contamination warning entry			1			1							
S116_HUMAN	S100 calcium-binding protein A16									1	1			
C14A_HUMAN	Dual specificity protein phosphatase CDC14A											1	1	

Second, results have been obtained using heterogeneous samples in particular cell lines, whole tissue biopsies, and epithelial cells purified from surgical specimens. Results obtained in cell lines do not allow accurate comparison between normal and cancer cells, and the presence/absence of proteins of interest has to be confirmed in biopsies. For example, when 2-D PAGE protein patterns of normal human

colonic crypts were compared with the CRC cell line LIM 1863, the proteins spots from normal crypts matched only 75–80% of the cell line spots and the relative expression levels of a large number of proteins differed [5]. When clinical biopsies are examined, results of phenotypic comparisons depend on the type of samples examined (e.g., heterogeneous whole tissue biopsies with inflammatory cells,

Table 5. Results of proteomics research in human CRC. Most cited up-regulated (orange), down-regulated (green), or mentioned (yellow) proteins out of 12 large-scale proteomics studies

Accession Number	Swiss-Prot	Entry Name	Name	SDS-PAGE	Free-flow electrophoresis	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE
				Simpson et al 2000 [29]	Simpson et al 2001 [31]	Demalte et al 2003 [33]	Reymond et al 1999 [9]	Medjahed et al 2003 [32]	Lawrie et al 2001 [30]	Friedman et al 2004 [35]	Sierum et al 2003 [34]
AF041260	O75363	ACTB_HUMAN	mRNA fragment encoding cytoplasmic actin								
AA232508	Q9Y397	ACTG_HUMAN	Actin, Cytoplasmic 2 (Gamma-Actin)								
AA007218	Q96QY8	GTP_HUMAN	Glutathione S-transferase M3 (brain)								
X77584	G01119	THIO_HUMAN	Thioredoxin (ATL-derived factor) (ADF)								
X75821	Q04984	ANX3_HUMAN	Annexin III (Lipocortin III)								
U22055	Q96AG0	ANX4_HUMAN	Annexin IV (Lipocortin IV)								
U14631	P80365	K1CR_HUMAN	Keratin, type I cytoskeletal 18 (cytokeratin 18) (K18) (CK18)								
M92439	P42704	ATPB_HUMAN	ATP5B ATP synthase								
AF029082	P31947	PPIA_HUMAN	Peptidyl-Prolyl Cis-Trans Isomerase A (Cyclophilin A)								
M86400	P29312	PDA3_HUMAN	Probable protein disulfide isomerase ER-60 precursor								
L76465	P15428	TPIS_HUMAN	Triosephosphate isomerase (TIM)								
X57352	Q01628	TCTP_HUMAN	Translationally controlled tumor protein (TCTP)(p23)								
X03205	P10809	CH60_HUMAN	Heat shock protein 60								
M95787	Q01995	VINC_HUMAN	Vinculin								
Z25821	P42128	EF1G_HUMAN	Elongation factor 1-gamma mRNA								
U96132	Q99714	TBA1_HUMAN	Tubulin alpha-1 chain, brain specific								
AB028893	P04643	CATD_HUMAN	Cathepsin D (EC 3.4.23..5).(GENE:CTSD)Homo sapiens								
X79239	Q02546	EF11_HUMAN	Elongation factor 1-alpha, mRNA								
M13934	P06366	G3P2_HUMAN	Glyceraldehyde 3 phosphate dehydrogenase								
M60854	P17008	GR78_HUMAN	78 kDa glucose-regulated protein precursor (GRP 78)								
M13932	P08708	ENO4_HUMAN	Alpha enolase (2-phospho-D-glycerate hydrolyase)								
X69150	P25232	CRTC_HUMAN	Calreticulum precursor (CR55) (calregulin)								
L06498	P17075	COF1_HUMAN	Cofilin, non-muscle isoform								
L04483	P35265	HS7C_HUMAN	Heat shock Cognate 71 kDa protein								
M64716	P25111	HBB_HUMAN	Hemoglobin beta chain.(GENE: HBB) h.s								
X77770	P02383	K2C8_HUMAN	Keratin, type II cytoskeletal 8 (cytokeratin 8) (K 8) (CK 8)								
W52460	P14798	RM12_HUMAN	Mitochondrial 60s ribosomal protein L7/L12 precursor								
X55715	P23396	GR75_HUMAN	Mitochondrial stress-70 protein precursor (GRP 75)								
M77234	P49241	SODC_HUMAN	Superoxide Dismutase (Cu-Zn)								
M58458	P12750	KCRB_HUMAN	Creatine kinase								
D16992	P10680	143Z_HUMAN	14-3-3 protein zeta/delta (protein kinase C inhibitor protein-1)								
M77233	P23821	RL4_HUMAN	60s ribosomal protein L4 (L1)								
F16294	P05358	ANX1_HUMAN	Annexin I (lipocortin I) (calpactin I) (chromogardin 9) (p35)								
X61156	P08865	ROA1_HUMAN	DNA binding protein UPI, liver mRNA fragment								
L36055	Q13541	FABL_HUMAN	L-FABP liver fatty acid-binding protein 1								

Table 5. Continued

Accession Number	Swiss-Prot	Entry Name	Name	SDS-PAGE	Free-flow electrophoresis	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE
AA316619	P04645	PAB1_HUMAN	Poly(A) binding protein, mRNA								
L14599	Q15233	RL3_HUMAN	Ribosomal protein L 3								
X79234	P39026	RS19_HUMAN	S 19 ribosomal protein								
L06505	P30050	SAHH_HUMAN	S-Adenosylhomocysteine hydrolase AHCY								
X64707	P26373	TERA_HUMAN	Transitional endoplasmic reticulum ATPase (TER ATPase)								
X56932	P40429	GDIR_HUMAN	Rho GDP-dissociation inhibitor 1								
L25899	P39030	ACON_HUMAN	Aconitate hydratase, mitochondrial precursor (aconitase)								
X63527	P14118	AKA1_HUMAN	Alcohol dehydrogenase (NADP+) (aldehyde reductase)								
X89401	P46778	AAC4_HUMAN	Alpha-Actinin 4								
H59771	P23131	NPM_HUMAN	B23 nucleophosmin								
M94314	P38663	CATA_HUMAN	Catalase								
L19527	P08526	CD44_HUMAN	CD44 antigen (Phagocytic glycoprotein I)								
U14968	P46776	TCPZ_HUMAN	Chaperonin containing TCP1, subunit 6A(zeta1)								
U14969	P46779	COXA_HUMAN	Cytochrome c oxidase polypeptide Va, mitochondrial (EC 1.9.3.1) (GENE:COX5A)H.s								
L38941	P49207	COXB_HUMAN	Cytochrome C oxidase polypeptide VB precursor.COX5B								
U12465	P42766	EF1B_HUMAN	Elongation factor 1-beta (EF-1-beta)								
F19234	P18077	ER29_HUMAN	Endoplasmic reticulum protein ERp29 (ERp31) (ERp28).								
X66699	P12751	EZRI_HUMAN	Ezrin (p81) (cytovillin)(villin-2)								
D23660	P36578	ALFA_HUMAN	Fructose-Biphosphate Aldolase A (Muscle-Type Aldolase)								
U14966	P46777	DHE3_HUMAN	Glutamate dehydrogenase 1 precursor (GDH)								
X06705	P11518	HS9A_HUMAN	Heat shock Protein HSP 90-Alpha								
D14531	P32969	ROL_HUMAN	Heterogeneous nuclear ribonucleoprotein L (HNRPN L)								
U30255	P52209	K2C1_HUMAN	Keratin, type II cytoskeletal 1 (cytokeratin 1) (K1) (CK1)								
M81182	P28288	PDX1_HUMAN	Peroxiredoxin 1 (EC 1.11.1.-) (Thioredoxin peroxidase 2)								
U04627	P40939	O60506	PRM RNA binding protein Gry-rbp (GRY-RBP)mRNA								
X87949	P11021	PHB_HUMAN	Prohibitin								
M16660	P08238	PDI_HUMAN	Protein disulfide isomerase precursor (PDI)								
U79725	Q99795	SPCB_HUMAN	Spectrin beta chain, erythrocyte (Beta-I spectrin)								
D10511	P24752	PDX3_HUMAN	Thioredoxin-dependent peroxide reductase, mitochondrial								
M93040	P22303	TPM3_HUMAN	Tropomyosin alpha 3 chain (Tropomyosin 3)								
AF102542	Q95395	NDKA_HUMAN	Non-metastatic cells 1, protein (NM23A)expressed in								
W45148	P24666	I43S_HUMAN	14-3-3 PROTEIN SIGMA								
M17885	P05387	SDH2_HUMAN	SDS ribosomal protein L24(L28)								
M17886	P05386	APA1_HUMAN	Apolipoprotein A-I								
M17887	P05387	DHSA_HUMAN	SDH2 succinate dehydrogenase flavoprotein subunit								

Table 6. Reproducibility of transcriptome results using proteomics tools (a) and *vice versa* (b). Proteomics results have been reproduced in 67% of cases in transcriptomics studies. Results from transcriptomics studies are more difficult to reproduce using proteomics tools (25%). Only two gene products can be retrieved in three and more proteomics and transcriptomics studies (actin and creatin-kinase)

a. Transcriptomics > Proteomics

				TRANSCRIPTOMICS										PROTEOMICS									
				cDNA	Adapter-tagged	SAGE	cDNA	cDNA	cDNA	SAGE	SAGE	SAGE	SAGE	cDNA	Oligonucleotide Array	SDS-PAGE	Free-flow electrophoresis	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE	2-D PAGE
Accession	Number	Swiss-Prot	Entry Name	Name	Paired normal and cancer	Normal and cancer	Muc,adenoma and pr.tum	Paired n. and pr.tumors	Paired n. and pr.tumors	Paired tumor and pr.tum or adenoma	Cell lines	Norm.tumc and Pr.tum	Primary tumors and cell lines	Cell line	Paired tumor and pr.tum or adenoma	Paired tumor and pr.tum or adenoma	Cell lines	Cell lines	Cell lines	Purified epithelial cells	In silico	Purified epithelial cells	Human whole tissue
					Yanagawa et al 2001 [21]	Muro et al 2003 [27]	Buckhaults et al 2001 [24]	Birkenkamp et al 2002 [25]	Takemasa et al 2001 [22]	Lin et al 2002 [26]	Paré-McDemott 2000 [20]	Zhang et al 1997 [16]	SAGE-NET et al 1997 [19]	Zhang et al 1997 [18]	Williams et al 2003 [28]	Netterman et al 2001 [23]	Simpson et al 2000 [29]	Simpson et al 2001 [31]	Demalto et al 2003 [33]	Reynold et al 1999 [9]	Medjahed et al 2003 [32]	Lawrie et al 2001 [30]	Friedman et al 2004 [35]
																							Sterum et al 2003 [34]
AF041260	O75363	CAH2_HUMAN	Carbonic anhydrase II																				
AA232508	Q9Y397	CEA1_HUMAN	EST (BILIARY GLYCOPROTEIN)																				
AA007218	Q96QY8	BENE_HUMAN	BENE																				
X05014	G01119	CAH1_HUMAN	Carbonic anhydrase I																				
X75821	Q04984	GUAV_HUMAN	GCAP-II guanylate cyclase activator 2B; Uroguanylin; UGN																				
U22055	Q96AG0	IFM2_HUMAN	Interferon-inducible protein 1-Bd																				
U14631	P80365	FAB1_HUMAN	L-FABP liver fatty acid-binding protein 1																				
M92439	P42704	SAKH_HUMAN	S-Adenosylhomocysteine hydrolase AHCY																				
AF029082	P31947	BGH3_HUMAN	Transforming growth factor, beta-induced, 66 kDa																				
M86400	P29312	ACDS_HUMAN	Acyl-CoenzymeA dehydrogenase C-2 to C-3 short chain																				
L76465	P15428	ADH6_HUMAN	ADH gamma 2 subunit (aa-1-375)																				
X57352	Q01628	CC16_HUMAN	APC6 or CDC16Hs cell division cycle 16																				
X03205	P50991	TCPD_HUMAN	Chaperonin containing TCP-1, subunit 4 (delta)																				
M95787	Q01995	T4S3_HUMAN	Co-029;transmembrane 4 superfamily member 3																				
Z25821	P42126	CA11_HUMAN	COL1A1																				
U96132	Q99714	DRA_HUMAN	Colon mucosa- associated (DRA)																				
AB028893	P04643	KCRB_HUMAN	Creatine kinase																				
X79239	Q02546	TF3A_HUMAN	General transcription factor IIIA																				
M13934	P06366	GRO_HUMAN	GRO1 oncogene(melanoma growth stimulating activity, alpha)																				
M60854	P17008	GUAN_HUMAN	GUCA1B guanylate cyclase activator 1B/guanylin																				
M13932	P08708	Q95784	IgG Fc-binding protein																				
X69150	P25232	ITMC_HUMAN	Integral membrane protein 2C																				
LD6498	P17075	MK03_HUMAN	MAPK3																				
LD4483	P35265	FXV3_HUMAN	MATB protein																				
M64716	P25111	MDP1_HUMAN	MPDPA, MDP7 microsomal dipeptidase																				
X77770	P02383	MPI2_HUMAN	M-phase inducer phosphatase 2																				
W52460	P14798	ACTB_HUMAN	mRNA fragment encoding cytoplasmic actin																				
X57715	P33206	BRAC_HUMAN	BRCA1/2, breast cancer 1/2, early onset, protein																				
M77234	P45241	SPRC_HUMAN	Secreted protein, acidic, cysteine-rich (osteonectin)																				
M58458	P12750	SELP_HUMAN	Selenoprotein P																				
D16992	P10660	EST1_HUMAN	EST																				

TRANSCRIPTOMICS	PROTEOMICS
------------------------	-------------------

www.proteomics-journal.de

necrosis, blood, stool, etc.), so that standardized sample preparation procedures are critical for obtaining reproducible results. Several sample preparation methods have been described, in particular fluorescence-activated cell-sorting (FACS) [4], laser capture microdissection (LCM) [5–8], immunomagnetic beads separation [9], and cellular fractionation [10]. Unfortunately, these sample preparation procedures were barely applied in CRC. In our experience, the beads-based method is characterized by several advantages when compared with other cell purification procedures [11].

Third, gene expression patterns depend on the arrays technology platform. In transcriptomics studies any factors may affect the outcome of a microarray experiment, in particular technical, instrumental, computational, and interpretative factors. In fact, lack of reproducibility and accuracy is a major concern in microarray studies [12]. When cross-platform comparison was performed, reproducibility was insufficient: only four genes from a set of 185 common genes selected behaved consistently on three array platforms, and agreement of about 30% was found between two brands [13].

Fourth, in proteomics studies, 2-D PAGE or 2-D DIGE have well-known technological limitations. In CRC, even after epithelial cell enrichment using magnetic beads, the mean CV of repeated 2-D PAGE analysis with silver staining was found to lie between 20 and 28%. Only 47% (interrun) to 76% (intrarun) of spots could be matched within a triplicate experiment. Interindividual phenotypic variability was high. Thus, even under well-defined experimental conditions, 2-D PAGE parallel analysis of paired CRC samples is hampered by a significant variability [2].

Fifth, the methods applied for generating, formatting, storing, retrieving, and querying data are of outmost importance to assess methodological and biological variation in gene expression analysis. Unfortunately, due to the small sample size (number of patients), large number of variables examined at once, and absence of double or triple experiments (arrays and gels are expensive and samples are rare) statistical analysis is often not valid. In particular, assessing the reproducibility of a variable is necessary (e.g., using the intraclass correlation coefficient) for comparing multiple samples at once. The use of median values instead of mean values has been shown to improve data correction [14]. It has also been proposed to use housekeeping genes as endogenous controls [15]. A dedicated society, The Microarray Gene Expression Data (MGED) Society, has been formed to facilitate the sharing of gene expression data generated by functional genomics and proteomics experiments [16].

Finally, correlation between results of transcriptomics versus proteomics results is low. For CRC, there is no publication comparing mRNA and protein expression for a cohort of genes. However, extrapolation is reasonable from another epithelial cancer (lung adenocarcinoma), where such comparison has been performed. Only a subset of the proteins (17%) exhibited a significant correlation with mRNA abundance [17].

Obviously, many genes and factors found to be differentially regulated (both in transcriptomics and proteomics studies) do not play a causal role in CRC carcinogenesis. For example in the studies under investigation, at least 17 mRNAs encoding ribosomal proteins were identified to be dysregulated using cDNA arrays and 39 ESTs using SAGE technology. This broad dysfunction of protein synthesis, in particular of small molecules synthesis, has been reported not only in cancer but also in several other human diseases, where etiologies have been linked to mutations in genes of the translational control machinery [36].

However, some findings might be of particular interest in human CRC. For example, a small molecule group found to be dysregulated in human CRC is the 14–3–3 proteins family. 14–3–3 proteins are ubiquitous within all eukaryotic cells and participate in protein kinase signaling pathways. In particular, they are involved in phosphorylation-dependent protein–protein interactions that control progression through the cell cycle, initiation and maintenance of DNA damage checkpoints, activation of MAP kinases, prevention of apoptosis, and coordination of integrin signaling and cytoskeletal dynamics [37]. Given the prevalence of specific 14–3–3 isoforms expression in several human epithelial cancers [38, 39], these proteins may be involved in cancer tumorigenesis and particular isoforms may be useful as therapeutic targets in human CRC.

In summary, we propose a gene expression data warehouse in human CRC that is intended to help researchers active in the field to get an overview of the data available. However, reproducibility of results obtained in different studies was disappointing.

As a matter of fact, the development of some types of unified processes for generating information, formatting, storing, retrieving, and querying data, regardless of the technology used to generate it, is of central importance for building gene expression databases. Such unified processes have been proposed (see above) but have found only limited recognition so far. This is unfortunate because these methodological issues must be solved before trying to integrate experimental data from different sources into functional proteomics studies at the bench or *in silico*. The present meta-analysis of gene expression highlights the need for including a sufficient number of patients, analyzing only purified epithelial cells, using clinical standards (such as the international classification of disease, histopathology, and staging), assessing the variability of the technology applied, considering common standards for microarray data annotation and exchange, and finally for developing software implementing these standards and promoting the sharing of these high quality, well annotated data within the life sciences community.

In the absence of such unified processes, reproducibility of results within a laboratory will continue to be low, communication of results between different groups will stay difficult, and identification of diagnostic and therapeutic targets will remain a lottery. These problems are challenging the whole drug dis-

covery process from the very beginning, namely target identification, so that target validation, assay development, and prioritization of compounds remains a high risk endeavor.

We are very grateful to G. Nestler for her excellent technical assistance.

5 References

- [1] Friedman, D. B., Hill, S., Keller, J. W. et al., *Proteomics* 2004, 4, 793–811.
- [2] Ott, V., Guenther, K., Steinert, R. et al., *Pharmacogenomics J.* 2001, 1, 142–151.
- [3] Ji, H., Whitehead, R. H., Reid, G. E. et al., *Electrophoresis*, 1994, 15, 391–405.
- [4] Raymond, M. A., Sanchez, J. C., Schneider, C. et al., *Electrophoresis* 1997, 18, 622–624.
- [5] Kondo, T., Seike, M., Mori, Y. et al., *Proteomics* 2003, 3, 1758–1766.
- [6] Banks, R. E., Dunn, M. J., Forbes, M. A. et al., *Electrophoresis* 1999, 20, 689–700.
- [7] Craven, R. A., Banks, R. E., *Proteomics* 2001, 1, 1200–1204.
- [8] Craven, R. A., Totty, N., Harnden, P. et al., *Am. J. Pathol.* 2002, 160, 815–822.
- [9] Raymond, M. A., Sanchez, J. C., Hughes, G. J. et al., *Electrophoresis* 1997, 18, 2842–2848.
- [10] Szymczyk, P., Krajewska, W. M., Jakubik, J. et al., *Tumori* 1996, 82, 376–381.
- [11] Kellner, U., Steinert, R., Seibert, V. et al., *Pathol. Res. Pract.* 2004, 200, 155–163.
- [12] Shi, L., Tong, W., Goodsaid, F. et al., *Expert Rev. Mol. Diagn.* 2004, 4, 761–777.
- [13] Tan, P. K., Downey, T. J., Spitznagel, E. L. Jr. et al., *Nucleic Acids Res.* 2003, 31, 5676–5684.
- [14] Pellis, L., Franssen-van Hal, N. L., Burema, J., Keijer, J., *Physiol. Genomics* 2003, 16, 99–106.
- [15] Janssens, N., Janicot, M., Perera, T., Bakker, A., *Mol. Diagn.* 2004, 8, 107–113.
- [16] www.mged.org/, consulted on December 25, 2004.
- [17] Chen, G., Gharib, T. G., Huang, C. C. et al., *Mol. Cell. Proteomics* 2002, 1, 304–313.
- [18] Zhang, L., Zhou, W., Velculescu, V. E. et al., *Science* 1997, 276, 1268–1272.
- [19] www.sagenet.org/cancer/table1.htm, consulted on December 1, 2003.
- [20] Parle-McDermott, A., McWilliam, P., Tighe, O., Dunican, D., Croke, D. T., *Br. J. Cancer* 2000, 83, 725–728.
- [21] Yanagawa, R., Furukawa, Y., Tsunoda, T. et al., *Neoplasia* 2001, 3, 395–401.
- [22] Takemasa, I., Higuchi, H., Yamamoto, H. et al., *Biochem. Biophys. Res. Commun.* 2001, 285, 1244–1249.
- [23] Notterman, D. A., Alon, U., Sierk, A. J., Levine, A. J., *Cancer Res.* 2001, 61, 3124–3130.
- [24] Buckhaults, P., Rago, C., St. Croix, B. et al., *Cancer Res.* 2001, 61, 6996–7001.
- [25] Birkenkamp-Demtroder, K., Christensen, L. L., Olesen, S. H. et al., *Cancer Res.* 2002, 61, 4352–4363.
- [26] Lin, Y. M., Furukawa, Y., Tsunoda, T. et al., *Oncogene* 2002, 21, 4120–4128.
- [27] Muro, S., Takemasa, I., Oba, S. et al., *Genome Biol.* 2003, 4, R21.
- [28] Williams, N. S., Gaynor, R. B., Scoggin, S. et al., *Clin. Cancer Res.* 2003, 9, 931–946.
- [29] Simpson, R. J., Connolly, L. M., Eddes, J. S. et al., *Electrophoresis* 2000, 21, 1707–1732.
- [30] Lawrie, L. C., Curran, S., McLeod, H. L., Fothergill, J. E., Murray, G. I., *Mol. Pathol.* 2001, 54, 253–258.
- [31] Simpson, R. J., *Proteomics* 2001, 1, 807–818.
- [32] Medjahed, D., Luke, B. T., Tontesh, T. S. et al., *Proteomics* 2003, 3, 1445–1453.
- [33] Demalte, I., Swiss-2-D PAGE database: Search for DLD1_Human. <http://www.expasy.org>, Release 16 and updates up to 17-Nov-2003.
- [34] Stierum, R., Gaspari, M., Dommels, Y., et al., *Biochim. Biophys. Acta* 2003, 1650, 73–91.
- [35] Friedman, D. B., Hill, S., Keller, J. W. et al., *Proteomics* 2004, 4, 793–811.
- [36] Calkhoven, C. F., Muller, C., Leutz, A., *Trends Mol. Med.* 2002, 8, 577–583.
- [37] Wilker, E., Yaffe, M. B., *J. Mol. Cell Cardiol.* 2004, 37, 633–642 (review).
- [38] Qi, W., Liu, X., Qiao, D. et al., *Int. J. Cancer* 2005, 113, 359–363.
- [39] Moreira, J. M., Gromov, P., Celis, J. E., *Mol. Cell. Proteomics* 2004, 3, 410–419.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☒ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.